

第5章

路由器和网络互连

目前的现状是多种类型网络共存,并独立发展,每一种网络已经解决连接在该网络上的两个终端之间的通信功能,如以太网已经实现连接在以太网上的两个终端之间的 MAC 帧传输过程,问题是,任何一种网络都无法独立实现全球任何两个终端之间的通信功能,只有把多种有着不同的适用范围、不同的功能特性的网络互连在一起,才能实现全球任何两个终端之间的通信功能,这就需要一种新的独立于任何类型网络的端到端分组传输协议——IP,一种新的用于互连不同类型网络的设备——路由器。

5.1 网络互连

5.1.1 网络互连需要解决的问题

实现不同种类传输网络(异构网络)的互连就是实现两个连接在不同种类的传输网络上的终端之间的端到端通信。图 5.1 是实现以太网和公共交换电话网(Public Switched Telephone Network,PSTN)互连的互连网络结构,PSTN 是由 PSTN 交换机和互连 PSTN 交换机的物理链路构成的网络,通过呼叫连接建立过程建立两个结点之间的点对点语音信道。图 5.1 所示互连网络的核心是路由器,路由器为了实现以太网和 PSTN 的互连,必须具有以太网端口和 PSTN 端口,能够通过以太网端口实现和连接在以太网上的终端之间 MAC 帧的传输,同样,能够在 PSTN 端口和连接在 PSTN 上的终端之间建立语音信道,并通过语音信道实现数据传输。为了通过以太网实现路由器以太网端口和其他连接在以太网上的终端之间的 MAC 帧传输,以太网端口需要分配 MAC 地址。同样,为了通过呼叫连接建立过程在路由器 PSTN 端口和其他连接在 PSTN 上的终端之间建立语音信道,需要为 PSTN 端口分配电话号码。这样,必须经过路由器中继,才能实现连接在不同传输网络上的两个终端之间的通信过程,因此,图 5.1 中终端 A 至终端 B 的传输路径由两部分组成,一是终端 A 至路由器以太网端口的以太网交换路径,该交换路径根据以太网交换机中的转发表和路由器以太网端口的 MAC 地址(MAC R)确定。二是路由器 PSTN 端口和终端 B 之间的语音信道,该语音信道通过路由器 PSTN 端口和终端 B 之间的呼叫连接建立过程建立。实际的数据传输过程应该这样,终端 A 将需要传输给终端 B 的数据封装在以 MAC A 为源 MAC 地址、以 MAC R 为目的 MAC 地址的 MAC 帧中,通过以太网将该 MAC 帧传输给路由器以太网端口。路由器从 MAC 帧中分离出数据,确定数据的目的终端(终端 B)位于

PSTN 端口连接的 PSTN,通过呼叫连接建立过程建立路由器 PSTN 端口和终端 B 之间的点对点语音信道,然后通过点对点语音信道将数据传输给终端 B。虽然点对点语音信道不需要寻址,但为了检测出数据传输过程中发生的错误,仍然需要将数据封装成帧,帧中除了数据字段,还须有检错码字段。因此,路由器首先将数据封装成适合点对点语音信道传输的帧格式,然后通过点对点语音信道将帧传输给终端 B。但图 5.1 所示的互连网络结构实现这样的传输过程还存在一些问题。

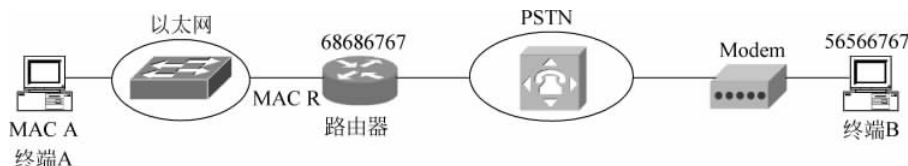


图 5.1 互连网络结构

1. 源终端路径选择问题

终端 A 向终端 B 传输数据前,需要获取终端 B 的地址,对于 PSTN,标识终端的地址是电话号码,如图 5.1 中终端 B 的电话号码 56566767。终端 A 即使获得了终端 B 的电话号码,如何确定终端 A 至终端 B 的传输路径? 对于图 5.1 所示的互连网络结构,终端 A 如何根据终端 B 的电话号码确定实现以太网和 PSTN 互连的路由器及路由器以太网端口的 MAC 地址?

2. 目的终端标识问题

终端 A 传输给终端 B 的数据封装在 MAC 帧中,路由器通过以太网端口接收到 MAC 帧后,需要通过呼叫连接建立过程建立路由器 PSTN 端口和终端 B 之间的语音信道,但 PSTN 的呼叫连接建立过程需要被叫和主叫的电话号码,路由器如何根据接收到的 MAC 帧及封装在 MAC 帧中的数据确定数据的目的终端及目的终端的电话号码?

3. 数据封装问题

适合以太网传输的数据封装形式是 MAC 帧,适合点对点语音信道传输的数据封装形式是点对点协议(Point-to-Point Protocol,PPP)帧,这是两种截然不同的封装形式,路由器根本无法根据 MAC 帧或 PPP 帧中包含的端到端传输的数据实现这两种封装形式的相互转换。

5.1.2 信件投递过程的启示

图 5.2 是将信件从南京投递到长沙的示意图,首先,寄信人用来传递信息的信纸被封装为信件,信封上写上寄信人和收信人地址,然后将信件提交给南京邮局,南京邮局根据收信人地址:长沙,确定信件的下一站:上海,由于南京至上海采用公路运输系统,因此,信件被封装为适合公路运输系统的形式:信袋,而且信袋上用车次 3536 表明运输该信袋的车辆及始站与终站。由于上海至长沙采用航空运输系统,上海首先从信袋中提取出信件,然后将其

封装成适合航空运输系统的形式：信箱，信箱上用航班号 AU765 表明运输该信箱的飞机及始站与终站。信件经过南京至上海的公路运输系统和上海至长沙的航空运输系统这两阶段的运输服务到达目的地：长沙。通过如图 5.2 所示的信件投递过程，可以得出以下启示：

- (1) 不同运输系统有着不同的封装信件的形式和标识始站与终站的方式；
- (2) 信件上收信人和寄信人地址是统一的，和实际提供运输服务的运输系统标识始站与终站的方式无关；
- (3) 信件是一种标准的封装形式，和实际提供运输服务的运输系统封装信件的形式无关；
- (4) 南京根据信件上的收信人地址确定下一站：上海，同样，上海也是根据信件上的收信人地址确定下一站：长沙，信件在南京至长沙的传输过程中是不变的；
- (5) 由实际的运输系统提供当前站至下一站的运输服务。

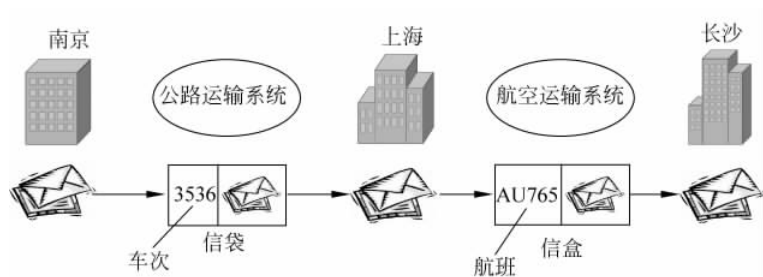


图 5.2 信件传输过程

5.1.3 端到端传输思路

根据图 5.2 所示的信件投递过程，可以引申出实现数据端到端传输的思路：

- (1) 定义一种和具体传输网络无关的、统一的终端地址格式：IP 地址。
- (2) 定义一种和具体传输网络无关的、统一的数据封装格式：IP 分组，并在 IP 分组中用 IP 地址标识数据的源和目的终端。
- (3) 如图 5.3 所示，端到端传输路径由终端、各种不同类型的传输网络和互连不同类型传输网络的路由器组成，终端和路由器称为跳，数据从源终端开始传输，首先传输给源终端至目的终端传输路径上的第一跳路由器，该路由器和源终端连接在同一个传输网络上，对应源终端为下一跳。数据端到端传输过程由多个传输阶段组成，每一个传输阶段实现数据从当前跳至和当前跳连接在同一传输网络的下一跳的传输过程，这种传输方式称为逐跳传输，每一跳通过 IP 分组中目的终端的 IP 地址确定下一跳。
- (4) 通过实际的传输网络实现 IP 分组当前跳至下一跳的传输过程，IP 分组经过实际传输网络传输时，须封装成实际传输网络对应的格式。

图 5.3 给出的数据端到端传输过程所涉及的步骤及功能如下：

- (1) 必须为终端 A、终端 B 及路由器分配统一的 IP 地址：IP A、IP B 和 IP R；
- (2) 终端 A 和路由器必须通过路由表给出用于根据终端 B 的 IP 地址 (IP B) 确定下一跳的信息；

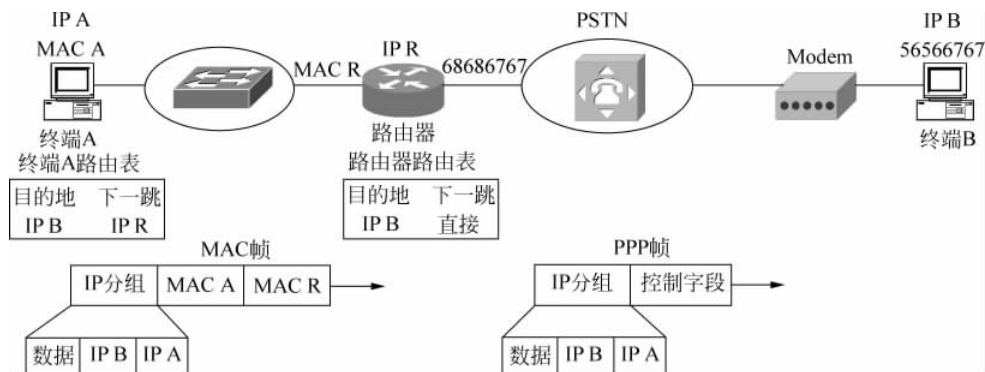


图 5.3 端到端数据传输过程

(3) 终端 A 将传输给终端 B 的数据封装成 IP 分组格式,并在 IP 分组中给出终端 A 和终端 B 的 IP 地址,端到端传输过程中,IP 分组是不变的;

(4) 终端 A 根据 IP 分组中终端 B 的 IP 地址确定下一跳: 路由器,并获得路由器的 IP 地址: IP R,终端 A 必须根据路由器的 IP 地址确定连接终端 A 和路由器的传输网络——以太网,获得路由器以太网端口的 MAC 地址,将 IP 分组封装成 MAC 帧,经过以太网将 MAC 帧传输给路由器;

(5) 路由器从 MAC 帧中分离出 IP 分组,根据 IP 分组中终端 B 的 IP 地址确定下一跳: 终端 B(路由器路由表中用“直接”表明下一跳就是目的终端自身),根据终端 B 的 IP 地址确定连接终端 B 和路由器的传输网络——PSTN,并获得终端 B 的电话号码,通过呼叫连接建立过程建立路由器和终端 B 之间的点对点语音信道,将 IP 分组封装成适合点对点语音信道传输的格式: PPP 帧,并经过点对点语音信道将 PPP 帧传输给终端 B;

(6) 终端 B 从 PPP 帧中分离出 IP 分组,再从 IP 分组中分离出终端 A 传输给它的数据。

5.1.4 IP 实现网络互连机制

1. 路由器和 IP 实现端到端传输的过程

从图 5.3 所示的端到端传输过程,可以得出网际协议(Internet Protocol, IP)实现网络互连的机制:

(1) 规定了统一的且与传输网络地址标识方式无关的 IP 地址格式,所有接入互连网络的终端必须分配一个唯一的 IP 地址,同时,由于每一个终端都和实际传输网络相连,还需具有实际传输网络相关的地址,如以太网的 MAC 地址,为了区分,将 IP 地址称为逻辑地址,将实际传输网络相关的地址称为物理地址。

(2) 规定了统一的且与传输网络数据封装格式无关的 IP 分组格式,端到端传输的数据必须封装成 IP 分组,每一跳通过 IP 分组携带的目的终端 IP 地址确定下一跳。

(3) 对应每一个目的终端,每一跳必须建立用于确定通往该目的终端的传输路径上的下一跳的信息,该信息被称为路由项,它主要由两部分组成: 目的终端 IP 地址和通往该目的终端的传输路径上的下一跳的 IP 地址。对应多个不同目的终端的路由项集合,称为路由表。

(4) 必须由单个传输网络连接当前跳和下一跳,能够根据下一跳 IP 地址确定连接当前跳和下一跳的传输网络,解析出下一跳的传输网络相关地址,即物理地址,能够将 IP 分组封装成传输网络要求的帧格式,并通过互连当前跳和下一跳的传输网络实现 IP 分组当前跳至下一跳的传输过程。

(5) IP 分组经过逐跳传输,实现源终端至目的终端的传输过程。

2. 路由表和路由项

实现互连网络端到端传输过程的关键有两点,一是源终端能够根据目的终端 IP 地址确定通往目的终端传输路径上的第一跳路由器,源终端至目的终端传输路径经过的路由器能够根据目的终端 IP 地址确定通往目的终端传输路径上的下一跳路由器。二是互连当前跳和下一跳的网络能够实现 IP 分组当前跳至下一跳的传输过程。如果互连当前跳和下一跳的网络是以太网,实现第二关键点需要完成以下三个步骤:①根据下一跳的 IP 地址解析出下一跳连接以太网接口的 MAC 地址;②将 IP 分组封装成以当前跳连接以太网接口的 MAC 地址为源 MAC 地址、下一跳连接以太网接口的 MAC 地址为目的地址的 MAC 帧;③经过以太网实现该 MAC 帧当前跳连接以太网接口至下一跳连接以太网接口的传输过程。前面有关以太网的章节已经详细讨论了完成第②步骤、第③步骤的方法和过程。实现第一关键点的关键是路由表,对应每一个目的终端地址,需要路由表给出通往该目的终端传输路径上的下一跳结点的 IP 地址。因此,路由表中的路由项格式是<目的终端 IP 地址,下一跳结点 IP 地址>,如果当前跳和目的终端连接在同一个网络,则当前跳至目的终端传输路径不存在其他路由器,下一跳结点 IP 地址用“直接”表示。如果当前跳通往一组目的终端的传输路径有着相同的下一跳结点,只需为这一组目的终端设置一项路由项<表示这一组目的终端的 IP 地址,下一跳结点 IP 地址>。

IP 分组端到端传输过程中,源终端将 IP 分组传输给第一跳路由器的过程,或者当前路由器根据目的 IP 地址和路由表确定下一跳路由器,并将 IP 分组传输给下一跳路由器的过程称为间接交付。源和目的终端位于同一个传输网络,或者路由器根据目的 IP 地址和路由表确定的下一跳是目的终端本身(目的 IP 地址匹配的路由项中的下一跳为“直接”),源终端或路由器通过直接连接的传输网络将 IP 分组传输给目的终端的过程称为直接交付。

5.1.5 数据报 IP 分组交换网络

为了简化互连网络端到端数据传输过程,在建立互连网络端到端传输路径时,可以将互连终端和路由器及路由器间互连的传输网络虚化成链路,将图 5.4(a)所示的互连网络结构简化为图 5.4(b)所示的由路由器互连多条链路而成的数据报分组交换网络,在图 5.4(b)所示的数据报分组交换网络中,路由器就是 IP 分组交换机,路由表就是转发表,每一个 IP 分组都是独立的数据报,路由器根据路由表和 IP 分组携带的目的终端 IP 地址实现 IP 分组的转发操作。由于可以将互连网络作为数据报 IP 分组交换网络进行分析、处理,因而常常用数据报 IP 分组交换网络(简称 IP 网络)来称呼互连网络。在 IP 网络中,传输网络的功能等同于逻辑链路,用于实现终端和 IP 分组交换机(路由器),及两个相邻 IP 分组交换机之间的 IP 分组传输,因而将传输网络的分组格式称为帧,将传输网络中的分组交换设备,如以太网交换机,称为链路层设备(第二层设备)。但 OSI 体系结构中定义的用于互连分组交换机的

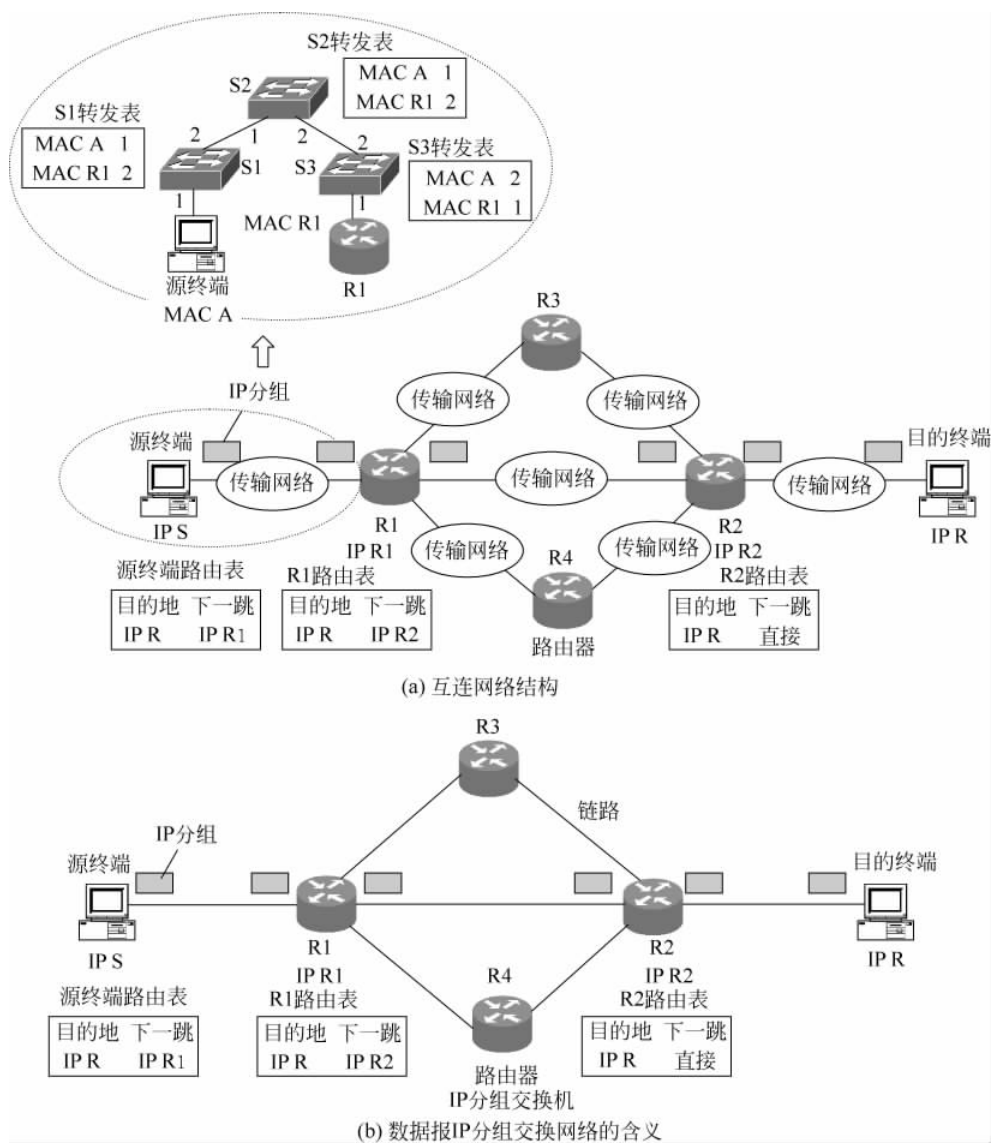


图 5.4 数据报 IP 分组交换网络

链路是点对点物理链路或广播物理链路(多点接入物理链路),因此,OSI 体系结构定义的网络应该是由终端、物理链路、分组交换机组成的,与图 5.4(a)所示的由终端、传输网络和用于互连传输网络的路由器组成的互连网络是不同的。对于图 5.4(a)所示的互连网络,端到端传输路径实际上由两层传输路径组成,一是 IP 层传输路径,由源终端、中间路由器和目的终端组成,如图 5.4(a)中源终端至目的终端传输路径:源终端→路由器 R1→路由器 R2→目的终端。二是传输网络中当前跳至下一跳的传输路径,如果连接源终端和路由器 R1 的传输网络是如图 5.4(a)中展开的交换式以太网,则源终端至路由器 R1 传输路径就是由源终端、中间以太网交换机和路由器 R1 组成的交换路径:源终端→以太网交换机 S1→以太网交换机 S2→以太网交换机 S3→路由器 R1。因此,和互连网络有关的内容由三部分组成,

它们分别是：IP、路由协议和 IP over X 技术。IP 详细规定了 IP 地址格式和 IP 分组格式。路由协议通过为每一个路由器建立路由表实现建立 IP 层传输路径的功能。IP over X(X 指特定的传输网络)技术实现根据下一跳 IP 地址解析出下一跳连接 X 传输网络的端口的物理地址,和 IP 分组经 X 传输网络从当前跳传输给下一跳的功能。

5.1.6 路由器结构

路由器从本质上讲是 IP 分组转发设备,根据 IP 分组首部中的目的终端 IP 地址完成 IP 分组从输入端口至输出端口的转发过程。图 5.5 是路由器功能结构,从功能上可以把路由器分成三部分：路由模块、线卡和交换模块。

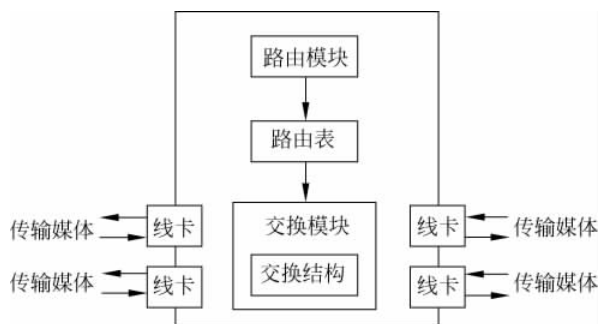


图 5.5 路由器结构

1. 路由模块

路由模块负责运行路由协议,生成路由表,在路由表中给出到达互连网络中任何一个终端的传输路径,当然,由于 IP 分组是逐跳转发,路由器的路由表中只需给出通往互连网络中某个终端的传输路径上的下一跳路由器的地址。由于生成路由表的过程比较复杂,因此,路由模块的核心部件通常是 CPU,大部分功能由软件实现。除了生成路由表,路由模块也承担一些其他的管理功能。

2. 线卡

线卡负责连接外部传输媒体,并通过传输媒体连接传输网络,如连接以太网的线卡通过双绞线或光纤连接以太网交换机。线卡通过端口连接传输媒体,不同类型的传输媒体对应不同类型的端口,如连接 5 类双绞线的端口(俗称电端口)和连接光纤的端口(俗称光端口)。线卡除了实现和传输网络的物理连接,还需要按照所连接的传输网络的要求完成 IP 分组的封装和分离操作。封装操作将 IP 分组封装成适合通过传输网络传输的链路层帧格式,如以太网的 MAC 帧。分离操作和封装操作相反,从链路层帧中分离出 IP 分组。线卡进行接收操作时,从所连接的传输媒体接收到的物理层信号(如曼彻斯特编码流)中分离出链路层帧(如 MAC 帧),并从链路层帧中分离出 IP 分组。发送操作时,将 IP 分组封装成链路层帧(如 MAC 帧),将链路层帧转换成物理层信号(如曼彻斯特编码流)后,发送到传输媒体。线卡的每一个端口还需配置输入、输出队列,用于存储无法及时处理的 IP 分组。

3. 交换模块

当线卡从某个端口接收到的物理层信号中分离出 IP 分组,就将该 IP 分组发送给交换模块。交换模块用 IP 分组的目的终端 IP 地址检索路由表,找到输出端口,并把 IP 分组发送给输出端口所在的线卡。随着端口的传输速率越来越高,如 10Gb/s 的以太网端口,端口每秒接收、发送的 IP 分组数量越来越大,对于 10Gb/s 的以太网端口,在极端情况下(假定 IP 分组的长度为 46B,MAC 帧的长度为 64B),端口每秒接收、发送的 IP 分组数量 $=10 \times 10^9 / (64 \times 8) = 19.53 \text{M IP 分组/s}$ (19.53Mpps),当路由器多个端口都线速接收、发送 IP 分组时,交换模块的处理压力将变得很大,因此,通常用称为交换结构的专用硬件来完成 IP 分组从输入端口到输出端口的转发处理。由于存在从多个输入端口输入的 IP 分组需要从同一个输出端口输出的情况,即使交换结构能够支持所有端口线速接收、发送 IP 分组,输出端口也需要设置输出队列,用输出队列来临时存储那些无法及时输出的 IP 分组。

路由器是实现不同类型的传输网络互连的关键设备,它一方面通过路由模块建立到达任何终端的传输路径,另一方面,在确定下一跳结点的 IP 地址后,完成下一跳结点 IP 地址到下一跳结点所连接的传输网络所对应的链路层地址的转换,并将 IP 分组封装成传输网络要求的链路层帧格式,通过传输网络传输给下一跳结点。

5.2 网际协议

网际协议(Internet Protocol,IP)是实现连接在不同类型传输网络上的终端之间通信功能的基础,用于定义独立于传输网络的 IP 地址和 IP 分组格式。

5.2.1 IP 地址分类

在深入讨论 IP 地址前,需要说明一下,IP 地址不是终端或路由器的标识符,而是终端或路由器接口的标识符,就像地址不是房子的标识符,而只是门牌号一样。一栋房子如果有多个门,则有多个不同的门牌号,也就有多个不同的地址,但以这些地址为收信人地址的信件都能投递给该房子的主人。同样,终端或路由器允许有多个接口,每一个接口都有独立的标识符——IP 地址,但以这些 IP 地址为目的地址的 IP 分组都能到达该终端或路由器。接口是终端或路由器连接网络的地方,多数情况下,终端或路由器的每一个端口都连接独立的网络,这种情况下,接口就是端口。但在一些特殊情况下,一个端口可能同时连接多个不同的网络,或是多个端口连接同一网络,因而,一个端口可能对应多个不同的接口,或是多个端口对应同一个接口。下面章节将针对具体应用,对这些特殊情况进行讨论。由于每一个 IP 地址指向唯一的终端或路由器,因此,从这一点上讲,IP 地址确实有终端或路由器标识符的作用。

1. IP 地址分类方法

图 5.6 给出了 IP 地址的分类方法。一般情况所指的 IP 地址是指 IPv4 所定义的 IP 地址,它由 32 位二进制数组成,为了表示方便,将 32 位二进制数分成 4 个 8 位二进制数,每个 8 位二进制数单独用十进制表示(0~255),4 个用十进制表示的 8 位二进制数用点分隔,如 32 位二

进制数表示的 IP 地址：01011101 10100101 11011011 11001001，表示成 93.165.219.201。

	1	2	3	4	
A	0	网络号	主机号		0.0.0.0~127.255.255.255
B	10	网络号	主机号		128.0.0.0~191.255.255.255
C	110	网络号	主机号		192.0.0.0~223.255.255.255
D	1110	组播地址			224.0.0.0~239.255.255.255
E	11110	保留			240.0.0.0~247.255.255.255

图 5.6 IP 地址分类方法

IP 是实现网络互连的协议，因此，用来标识互联网中终端设备的每一个 IP 地址由两部分组成：网络号和主机号。最高位为 0，表示是 A 类地址，用 7 位二进制数标识网络号，24 位二进制数标识主机号，A 类地址中网络号全 0 和全 1 的 IP 地址有特别用途，不能作为普通地址使用。0.0.0.0 表示 IP 地址无法确定，终端没有分配 IP 地址前，可以用 0.0.0.0 作为 IP 分组的源地址。127.X.X.X 是环回地址。所有类型的 IP 地址中，主机号全 0 和全 1 的 IP 地址也有特别用途，也不能作为普通地址使用。如网络号为 5 的 A 类 IP 地址的范围为 5.0.0.0~5.255.255.255，但 IP 地址 5.0.0.0 用于表示网络号为 5 的网络地址，而 IP 地址 5.255.255.255 作为在网络号为 5 的网络内广播的广播地址。A 类地址的范围是 0.0.0.0~127.255.255.255，但实际能用的网络号是 1~126，每一个网络号下允许使用的主机号= $2^{24}-2$ ，由此可以看出，A 类地址适用于大型网络。

最高 2 位为 10，表示 B 类地址，用 14 位二进制数标识网络号，用 16 位二进制数标识主机号，能够标识的网络号为 2^{14} ，每一个网络号下允许使用的主机号= $2^{16}-2$ 。B 类地址的范围是 128.0.0.0~191.255.255.255，适用于大、中型网络。

最高 3 位为 110，表示是 C 类地址，用 21 位二进制数表示网络号，8 位二进制数表示主机号，能够标识的网络号为 2^{21} ，每一个网络号下能够标识的主机号= 2^8-2 。很显然，C 类地址只适用于小型网络。实际应用中并不使用 B 类和 C 类地址中网络号全 0 的 IP 地址。

A、B、C 三类地址都称为单播地址，用于唯一标识 IP 网络中的某个终端，但任何网络内都有一个主机号全 1 的地址作为该网络内的广播地址，这种广播地址不能用于标识网络内的终端，只能在传输 IP 分组时作为目的地址，表明接收方是网络内的所有终端。

每一个单播 IP 地址具有唯一的网络号，因此，对应唯一的网络地址，根据单播 IP 地址求出对应的网络地址的过程如下：根据该 IP 地址的最高字节值确定该 IP 地址的类型，根据类型确定主机号字段位数，清零主机号字段得到的结果就是该 IP 地址对应的网络地址。如 IP 地址 193.1.2.7 对应的网络地址为 193.1.2.0。

最高 4 位为 1110，表示是组播地址，用 28 位二进制数标识组播组，同一个组播组内的终端可以任意分布在 Internet 中，因此，组播组是不受网络范围影响的。有些组播地址有特殊用途，称为著名组播地址，下面就是一些常用的著名组播地址，这些组播地址表明接收端是同一网络内的特定结点。

224.0.0.1 表示网络中所有支持组播的终端和路由器。

224.0.0.2 表示网络中所有支持组播的路由器。

224.0.0.4 DVMRP 路由器。

224.0.0.5 表示网络中所有运行 OSPF 进程的路由器。

224.0.0.9 表示网络中所有运行 RIP 进程的路由器。

最高 5 位为 11110,表示是 E 类地址,目前没有定义。

32 位全 1 的 IP 地址称为受限广播地址,以受限广播地址为目的地址的 IP 分组在当前网络内广播,所有路由器均不转发以受限广播地址为目的地址的 IP 分组。特定网络内主机号全 1 的 IP 地址称为直接广播地址,也称为定向广播地址,以直接广播地址为目的地址的 IP 分组在由网络号指定的特定网络内广播,所有没有和该特定网络直接相连的路由器正常转发该 IP 分组。

2. 互连网络 IP 地址配置原则

互连网络配置 IP 地址的原则如下:

(1) 连接在同一传输网络上的终端必须配置具有相同网络号的 IP 地址,如连接在以太网上的终端 A 和终端 C;

(2) 每一个传输网络都有一个网络地址,如图 5.7 中以太网配置的网络地址 192.1.1.0 和 PSTN 配置的网络地址 192.1.2.0;

(3) 路由器的每一个接口都需配置 IP 地址,该 IP 地址对应的网络地址必须和分配给该接口所连的传输网络的网络地址相同,如图 5.7 中连接以太网接口配置的 IP 地址 192.1.1.254,其网络地址为 192.1.1.0,和以太网配置的网络地址相同。

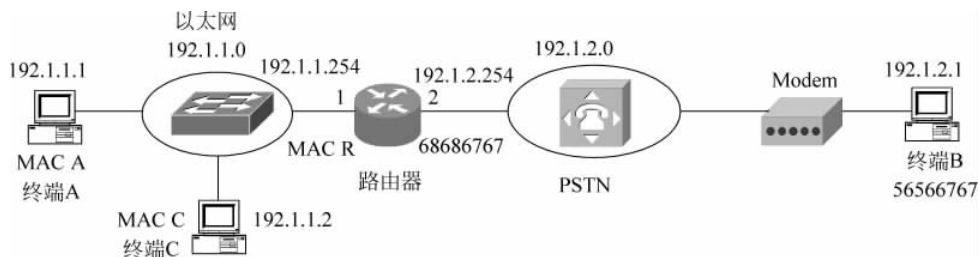


图 5.7 IP 地址配置

如果一个物理以太网被划分为多个 VLAN,则每一个 VLAN 就是一个独立的传输网络,不同 VLAN 须配置不同的网络地址,需要用路由器实现多个 VLAN 的互连。

5.2.2 IP 地址分层分类的原因和缺陷

1. 根据考号寻找考场的启示

图 5.8 假定考号由 6 位十进制数组成,其中高 3 位是考场号,低 3 位是座位号,同一考场的考号具有相同的考场号。每一个考场的考场号随机分配。用 751XXX 表示该考场的考场号是 751,座位号包括该考场内的全部座位号,因此,可用 751XXX 表示考场号是 751 的考场。

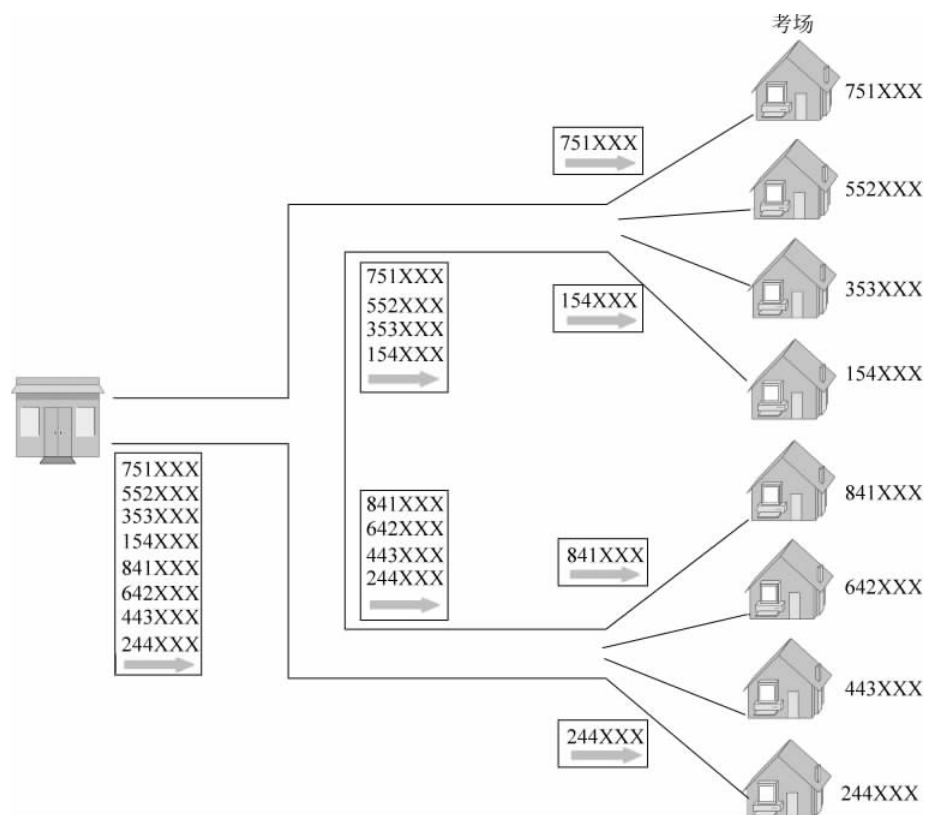


图 5.8 考场分布及引导方式

图中的考场指示方式必须保证能够将所有考号属于这 8 个考场的考生正确引导到考场指定的考场,每一个考生,在每一个路口,用考号的高 3 位比较路牌中各项的考场号,一旦考号中的考场号和路牌中某项的考场号相同,表示考号和该项匹配,考生将沿着该项给出的方向继续前进,直到正确到达考场指定考场,然后,再在考场内根据座位号找到正确的座位。由于考号分为两层,因此,路口路牌中的每一项只需给出特定考场的考场号及该考场的前进方向,无须为每一个考号设置前进方向。

两层结构减少了路牌中的项数,但当该单位设置多个考场时,路口路牌中的项数仍然偏多,有什么方式可以在两层结构不变的前提下,减少路牌中的项数?

如果考场号的分配方式如图 5.9 所示,符合以下分配原则:

- (1) 最高位为 6 的考场号只分配给分布在该单位的考场,且该单位所有考场的考场号的最高位一定为 6;
- (2) 分配给同一个区域中考场的考场号的次高位号必须是相同的,且分配给不同区域中考场的考场号的次高位必须不同。

根据上述分配原则分配考场号,可以进一步减少路口路牌的项数,但考生在不同路口用于匹配路牌中每一项的考号的位数是变化的,如单位入口用于匹配路牌中每一项的是考号的最高位,丁字路口用于匹配路牌中每一项的是考号的高 2 位,因此,必须在路口路牌中给出用于匹配路牌中每一项的考号位数。

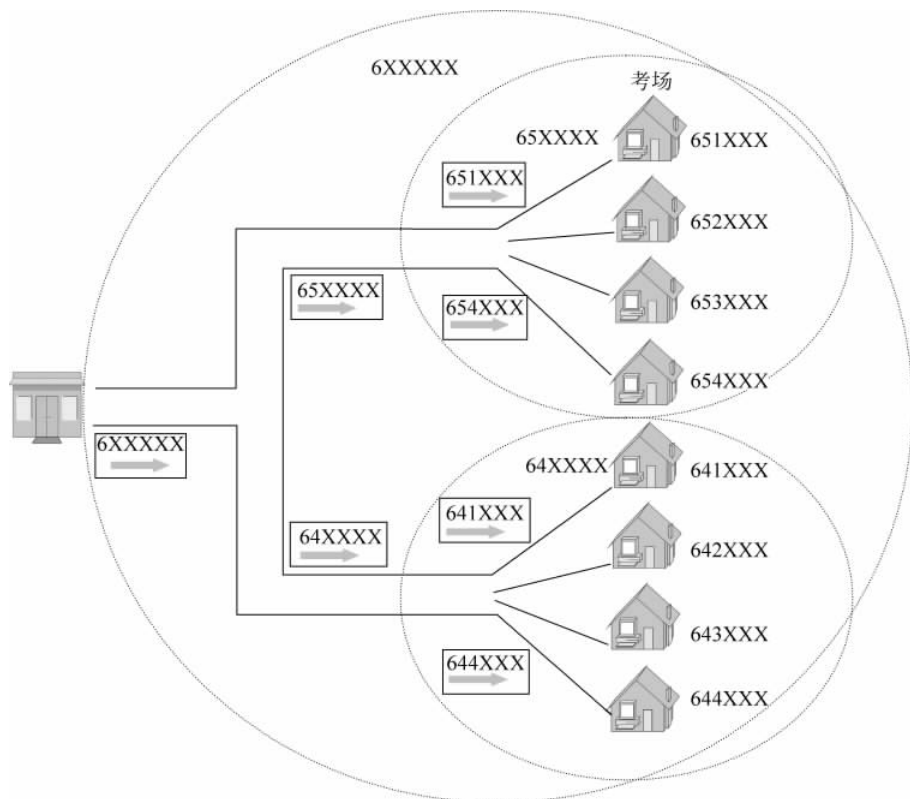


图 5.9 特定考场号分配规则下考场引导方式

根据两层考号寻找考号对应的考场座位的过程给出以下启示：

(1) 考生首先根据考号中考场号找到考场,然后,在考场寻找座位,因此,路口路牌只需指出通往每一个考场的路径;

(2) 通过限制考场号的分配,使得路口路牌可以用考场号的最高位或高 2 位指定分布在单位内的所有考场,或分布在单位内某个区域中的所有考场,以此减少路口路牌中的项数。

2. IP 地址分层分类的原因

IP 地址分层的目的也是希望用一项路由项指出通往该网络内所有终端的传输路径,如图 5.10 所示。需要说明的是:一般情况下,路由项格式是<目的网络,下一跳路由器 IP 地址>,目的网络字段给出目的终端所在网络的网络地址,这里为了讨论问题方便,假定图 5.10 中的路由器都直接用点对点物理链路相连,因此,路由项中只需给出转发端口就可确定通往目的终端的传输路径上的下一跳路由器,路由项格式变为<目的网络,转发端口>。但如果连接两个路由器的是类似以太网这样的传输网络,仅知道转发端口并不能确定通往目的终端的传输路径上的下一跳路由器,必须给出下一跳路由器的 IP 地址。

IP 地址为 192.1.1.1 的终端寻找通往 IP 地址为 192.2.1.1 的终端的传输路径的过程和图 5.8 中考生从单位入口开始根据考号寻找对应的考场座位的过程十分相似,每一个路

由器就像路口,而路由表就像路口路牌,转发端口就是前进方向,路由器根据目的终端的 IP 地址确定 IP 分组转发端口的过程如下:

(1) 求出目的终端 IP 地址对应的网络地址 N;

(2) 用 N 逐项比较路由表中每一项路由项的目的网络字段,如果和其中一项的目的网络字段值相同,表示目的终端的 IP 地址和该路由项匹配,通过该路由项指定的转发端口输出 IP 分组;

对于路由器 1,首先求出目的终端 IP 地址 192.2.1.1 对应的网络地址 192.2.1.0,然后逐项比较路由表中每一项路由项的目的网络字段值,结果和路由项<192.2.1.0,端口 2>的目的网络字段值相同,通过端口 2 输出该 IP 分组。两个终端之间传输路径上的所有路由器依次转发,最终将 IP 分组送达 IP 地址为 192.2.1.1 的终端。

IP 地址分类的原因是不同单位的网络规模是不同的,有些单位的网络规模很大,可以采用 B 类,甚至 A 类地址,有些单位的网络规模较小,可以采用 C 类地址,使得 IP 地址分配更贴近实际需要。

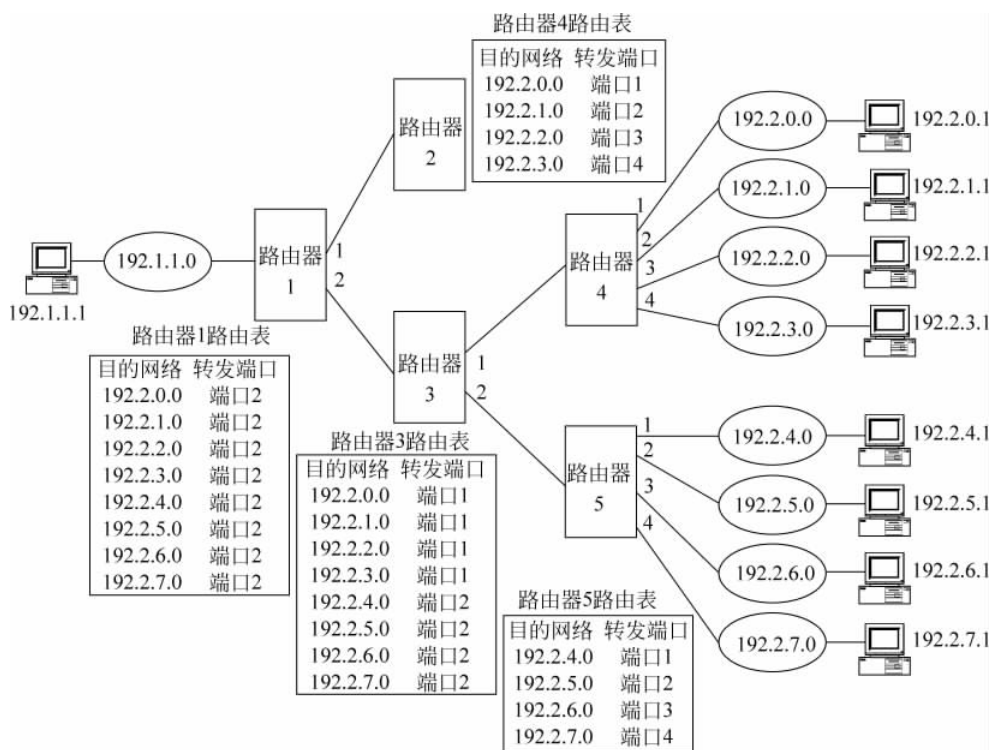


图 5.10 固定地址结构的路由方式

3. 固定地址结构的缺陷

图 5.10 所示的路由方式中,路由表必须为每一个网络设置一项路由项,这将使得路由表中的路由项数目非常庞大,由于大量网络之间的传输路径都需要经过核心路由器,因此核心路由器路由项数目庞大的问题尤为严重。路由项数目庞大,一是占用大量的存储空间,二是使得路由器根据目的终端 IP 地址确定下一跳路由器的转发处理变得耗时。

当初设计 IP 地址时,将 IP 单播地址分成 A、B、C 三类的目的是为了使 IP 地址能够适应不同规模的网络,地址分配能够更加贴近实际需要,避免浪费。但实际应用中,IP 地址浪费的问题依然十分严重,尽管将地址分成 A、B、C 三类,但许多网络规模介于两类地址之间,如具有 1000 个终端的网络,C 类地址不够,B 类地址又很浪费,因此,20 世纪 90 年代中期就出现 IP 地址短缺问题,而出现 IP 地址短缺问题的主要原因是大量 A 类和 B 类地址空间被浪费。

解决上述问题的关键是能够动态改变 IP 地址中网络号的位数,如图 5.9 所示的根据考号寻找考场座位的过程,如果不同路由器中和同一目的终端 IP 地址匹配的路由项的网络地址中的网络号的二进制位数是可变的,同样可以通过限制网络的网络号分配,用网络号的最高若干位表示网络号连续的一组网络号。更进一步,如果网络号的二进制位数不再固定,可以动态设置,可以根据实际网络规模申请主机号的位数,如 1000 台主机,可以申请一个 10 位主机号、22 位网络号的 IP 地址块,这样,传统的分类不再存在,网络号和主机号位数根据实际情况动态设置,这种思路就是无分类编址(Classless InterDomain Routing, CIDR, 直译是无分类域间路由)。

5.2.3 无分类编址

1. 无分类编址机制

无分类编址方式下,32 位 IP 地址中标识网络号和主机号的二进制位数是可变的,这样做,消除了 IP 地址的分类,也解决了因为分类带来的种种问题。但必须提出一种用于指明 IP 地址中作为网络号的二进制位数的方法。无分类编址通过子网掩码(更确切的名称应该是网络掩码,但子网掩码已经成为习惯称呼)指明 IP 地址中作为网络号的二进制位数。子网掩码也是一个 32 位的二进制数,和 IP 地址的表示方法一样,也用 4 个点分隔的十进制数表示,每个十进制数表示 8 位二进制数,如 255. 0. 0. 0,展开成二进制表示为 11111111 00000000 00000000 00000000。子网掩码中为 1 的二进制数对应 IP 地址中作为网络号的二进制数。5. 1. 1. 2/255. 0. 0. 0 表示 IP 地址是 5. 1. 1. 2,对应的子网掩码是 255. 0. 0. 0,如果将子网掩码展开成二进制表示,则只有高 8 位二进制数为 1,其余为 0,这就意味着 IP 地址的高 8 位为网络号,低 24 位为主机号。同样,5. 1. 1. 2/255. 255. 255. 0 表示 IP 地址的高 24 位为网络号,低 8 位为主机号。目前还有一种更直接的表示方式是直接给出 IP 地址中作为网络号的二进制位数,如 5. 0. 0. 0/8、5. 1. 0. 0/16、192. 2. 0. 0/21 等。更简单的表示方式是省略 IP 地址中低位连续的 0,如 5. 0. 0. 0/8 可以表示成 5/8,5. 1. 0. 0/16 可以表示成 5. 1/16。

图 5.9 中 6XXXXX 并不是表示一个 1 位考场号,5 位座位号的超大考场,而是考场号最高位为 6 的一组考场号的表示方式,同样,用 N 位网络前缀表示一组最高 N 位相同的连续网络号,网络前缀的表示方式和前面表示网络号的方式相同,可以用子网掩码或数字指定 32 位 IP 地址中网络前缀的位数,但网络前缀和网络号的含义不同,它只是用来表示具有相同网络前缀的一组 IP 地址,这一组 IP 地址称为 CIDR 地址块,可能由分配给不同网络的 IP 地址组成。〈网络前缀,主机号〉的 IP 地址结构完全取消了原先定义的 A、B、C 三类 IP 地址的概念,因而被称为无分类编址。 N 位网络前缀的 CIDR 地址块可以分配给单个网络,这

种情况下, N 位网络前缀就是该网络的网络号。也可以分配给多个网络, 这种情况下, N 位网络前缀只是用来确定 CIDR 地址块的 IP 地址范围。

2. 聚合路由项

图 5.10 中, 路由器 1 对应 8 个网络的路由项有这样的特点:

① 8 个网络的网络号是连续的, 所包含的 IP 地址范围为 192.2.0.0~192.2.7.255。这样范围内的 IP 地址的最高 21 位是相同的, 而且地址范围包含了低 11 位 ($21+11=32$) 的全部 2^{11} 种组合。具有这样特性的 IP 地址块可以表示为网络前缀为 21 位的 CIDR 地址块, 如果将 <网络前缀, 主机号> 的 IP 地址结构中主机号字段为 0 的地址称为网络前缀地址的话, 通过用最高 21 位为 1 的子网掩码和地址块中任何一个 IP 地址相与后获得的结果作为该 CIDR 地址块的网络前缀地址。以下是一个网络前缀地址的计算实例。

```

      11000000 00000010 00000001 00000001  192.2.1.1
& & 11111111 11111111 11111000 00000000  255.255.248.0
-----

```

```

      11000000 00000010 00000000 00000000  192.2.0.0

```

因此, 网络前缀地址 192.2.0.0/21 包含的 IP 地址范围等于图 5.10 中 8 个网络所包含的 IP 地址范围。

② 8 个网络对应的路由项有着相同的转发端口, 即有着相同的下一跳。

具有上述两个特点的 8 项路由项可以聚合为 1 项路由项 <192.2.0.0/21, 端口 2>。同样, 路由器 3 中网络 192.2.0.0、192.2.1.0、192.2.2.0 和 192.2.3.0 与网络 192.2.4.0、192.2.5.0、192.2.6.0 和 192.2.7.0 所包含的 IP 地址块构成网络前缀为 22 位的 CIDR 地址块, 这种地址块的网络前缀地址用最高 22 位为 1 的子网掩码和地址块中任何一个 IP 地址相与后获得, 如计算网络 192.2.4.0、192.2.5.0、192.2.6.0 和 192.2.7.0 所包含的 CIDR 地址块对应的网络前缀地址的过程如下:

```

      11000000 00000010 00000101 00000111  192.2.5.7
& & 11111111 11111111 11111100 00000000  255.255.252.0
-----

```

```

      11000000 00000010 00000100 00000000  192.2.4.0

```

网络前缀地址 192.2.4.0/22 包含的 IP 地址范围等于网络 192.2.4.0、192.2.5.0、192.2.6.0 和 192.2.7.0 包含的 IP 地址范围。同样, 网络前缀地址 192.2.0.0/22 包含的 IP 地址范围等于网络 192.2.0.0、192.2.1.0、192.2.2.0 和 192.2.3.0 包含的 IP 地址范围, 且这两组网络对应的路由项有着相同的转发端口, 因此, 每一组网络对应的 4 项路由项可以聚合为 1 项路由项 <192.2.0.0/22, 端口 1> 和 <192.2.4.0/22, 端口 2>。以此可以得出图 5.11 所示的采用无分类编址后的各路由器中的路由项。和图 5.10 比较, 图 5.11 中的路由项数目大幅减少。路由项聚合后得出的由 N 位网络前缀指定的 CIDR 地址块是分配给一组网络号连续且高 N 位相同的网络的 IP 地址集合, 只是通往这一组网络的传输路径有着相同的下一跳。

采用无分类编址后, 路由器根据 IP 分组的目的终端 IP 地址确定下一跳路由器的过程如下:

(1) 根据路由项的网络前缀位数求出目的 IP 地址的网络前缀地址, 然后和路由项目的网络字段值比较, 如果相同, 表示该目的 IP 地址和该路由项匹配, 通过路由项指定的转发

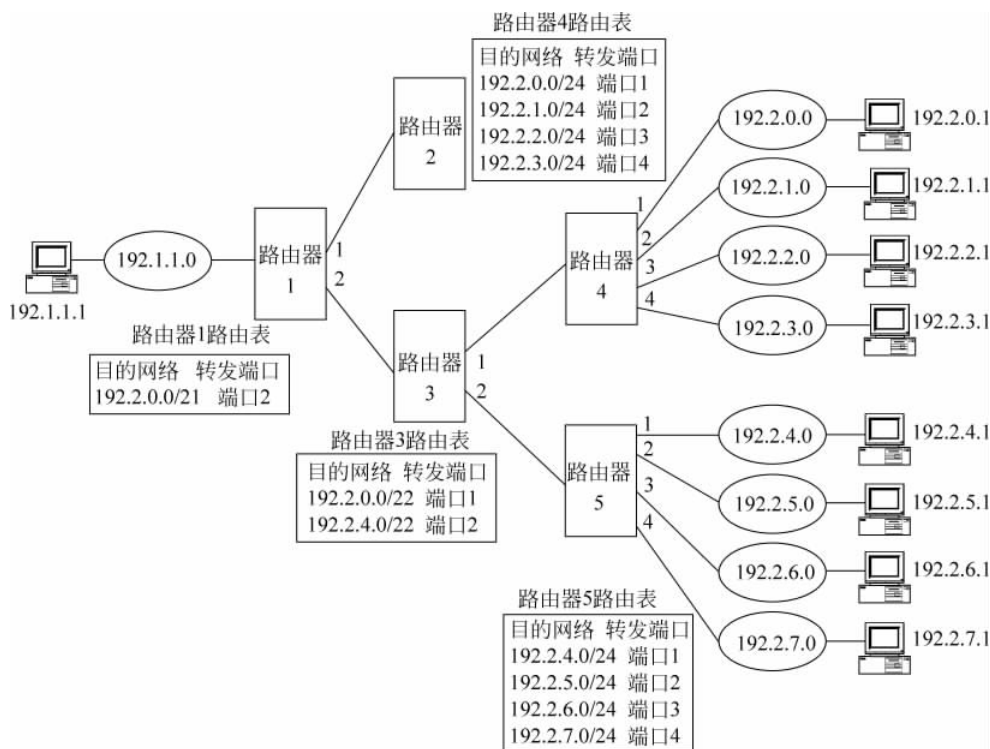


图 5.11 采用无分类编址后的路由项表示

端口输出该 IP 分组；

(2) 对路由表中的每一项路由项进行上述操作，直到找到匹配的路由项，或报错。

如果 IP 分组中的目的 IP 地址为 192.2.5.1，路由器 3 进行如下操作：

对于第一项路由项，由于网络前缀为 22 位，用最高 22 位为 1 的子网掩码 (255.255.252.0) 和 192.2.5.1 相与后获得对应的网络前缀地址。

```

      11000000 00000010 00000101 00000001   192.2.5.1
&& 11111111 11111111 11111100 00000000   255.255.252.0
-----
      11000000 00000010 00000100 00000000   192.2.4.0
  
```

由于目的 IP 地址 192.2.5.1 对应的网络前缀地址为 192.2.4.0/22，和路由项中的目的网络字段值 192.2.0.0/22 不同，因此，IP 分组的目的 IP 地址和第一项路由项不匹配。由于第二项路由项的网络前缀也是 22 位，目的 IP 地址对应的网络前缀地址同样为 192.2.4.0/22，它和第二项路由项中的目的网络字段值相同，因此，通过端口 2 输出该 IP 分组。

3. 任意划分子网

采用无分类编址的另一个好处是可以任意划分子网，假定某个单位有 120 台计算机，这些计算机被分成 6 组，其中第 1 组分配 20 台计算机，第 2 组分配 12 台计算机，第 3 组分配 45 台计算机，第 4 组分配 27 台计算机，第 5 组分配 5 台计算机，第 6 组分配 11 台计算机，这 6 组计算机属于 6 个子网，如何分配 IP 地址才能使得路由表中的路由项最少？为了使路由表中的路由项最少，分配给这些计算机的 IP 地址必须是连续的，因此，用最低 8 位二进制数

不同的 256 个 IP 地址作为这些计算机的 IP 地址,如 CIDR 地址块 192.1.2.0/24。最后 8 位二进制数的分配规则如下:

00000000~00111111(0~63)分配给第 3 组 45 台计算机,网络地址为 192.1.2.0/26;
 01000000~01011111(64~95)分配给第 4 组 27 台计算机,网络地址为 192.1.2.64/27;
 01100000~01111111(96~127)分配给第 1 组 20 台计算机,网络地址为 192.1.2.96/27;
 10000000~10001111(128~143)分配给第 2 组 12 台计算机,网络地址为 192.1.2.128/28;
 10010000~10011111(144~159)分配给第 6 组 11 台计算机,网络地址为 192.1.2.144/28;
 10100000~10100111(160~167)分配给第 5 组 5 台计算机,网络地址为 192.1.2.160/29。

上述分配过程的思路如下,最多的是第 3 组有计算机 45 台,求出满足不等式 $2^N \geq 45 + 2$ 的最小 N ,得出主机字段需要 6 位二进制数,8 位二进制数的表示范围可以分成 4 个 6 位二进制数的表示范围,高 2 位分别为 00、01、10 和 11。高 2 位为 00 的 64 个地址分配给第 3 组。第 4 组和第 1 组的计算机台数分别为 27 和 20,主机字段需要 5 位二进制数,高 2 位为 01 的 64 个地址可以分成高 3 位分别是 010 和 011 的 2 组 32 个地址,分别分配给第 4 组和第 1 组。第 2 组和第 6 组的计算机台数分别为 12 和 11,主机字段需要 4 位二进制数,高 2 位为 10 的 64 个地址可以分成高 4 位分别是 1000、1001、1010 和 1011 的 4 组 16 个地址,高 4 位分别为 1000 和 1001 的 2 组 16 个地址分别分配给第 2 组和第 6 组。高 4 位为 1010 的 16 个地址可以分成高 5 位分别是 10100 和 10101 的 2 组 8 个地址,将高 5 位为 10100 的 8 个地址分配给第 5 组。

图 5.12 中的路由器 R1 只需给出 1 项路由项 $\langle 192.1.2.0/24, 192.1.1.1, 1 \rangle$,表明只要目的 IP 地址高 24 位等于 192.1.2 的 IP 分组均转发给 IP 地址为 192.1.1.1 的下一跳路由器 R2,路由器 R2 对每一个子网均需给出 1 项路由项,目的网络字段值给出的 CIDR 地址块必须包含分配给该子网的全部 IP 地址。

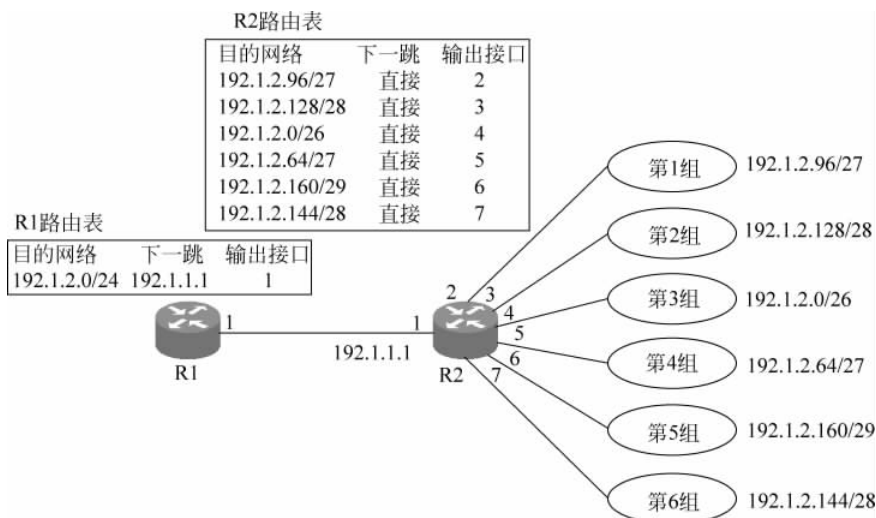


图 5.12 无分类编址任意划分子网过程

【例 5.1】 某网络的 IP 地址空间为 192.168.5.0/24,采用等长子网划分,子网掩码为 255.255.255.248,求子网数和每一个子网内可分配的 IP 地址数。

【解析】 根据子网掩码 255.255.255.248 得出网络号位数为 29 位,主机号位数为 3 位,192.168.5.0/24 IP 地址空间中,用 5 位($29-24=5$)作为子网号,求出子网数 $=2^5=32$,子网内可分配的 IP 地址数 $=2^3-2=6$ 。

【例 5.2】 某企业分配给人事部的 CIDR 地址块为 10.0.11.0/27,分配给企划部的 CIDR 地址块为 10.0.11.32/27,分配给市场部的 CIDR 地址块为 10.0.11.64/26,这三个 CIDR 地址块聚合后的 CIDR 地址块应是_____。

A. 10.0.11.0/25

B. 10.0.11.0/26

C. 10.0.11.64/25

D. 10.0.11.64/26

【解析】 答案是 A,CIDR 地址块 10.0.11.0/27 表示的 IP 地址集合是:10.0.11.0(IP 地址低 8 位为 000 00000)~10.0.11.31(IP 地址低 8 位为 000 11111),CIDR 地址块 10.0.11.32/27 表示的 IP 地址集合是:10.0.11.32(IP 地址低 8 位为 001 00000)~10.0.11.63(IP 地址低 8 位为 001 11111),两个 CIDR 地址块聚合后的 IP 地址集合是:10.0.11.0(IP 地址低 8 位为 00 000000)~10.0.11.63(IP 地址低 8 位为 00 111111),实际上就是 CIDR 地址块 10.0.11.0/26。CIDR 地址块 10.0.11.64/26 表示的 IP 地址集合是:10.0.11.64(IP 地址低 8 位为 01 000000)~10.0.11.127(IP 地址低 8 位为 01 111111),和 CIDR 地址块 10.0.11.0/26 聚合后的 IP 地址集合为 10.0.11.0(IP 地址低 8 位为 0 0000000)~10.0.11.127(IP 地址低 8 位为 0 1111111),实际上就是 CIDR 地址块 10.0.11.0/25。

4. 最长前缀匹配

图 5.12 中,假定 IP 分组的目的 IP 地址为 192.1.2.150,路由器必须找出通往 IP 地址为 192.1.2.150 的目的终端的传输路径,由于路由器 R1 在用 IP 地址 192.1.2.150 检索路由表时,判定该 IP 地址包含在由目的网络字段值 192.1.2.0/24 表示的 CIDR 地址块中(根据 24 位网络前缀计算出的目的 IP 地址 192.1.2.150 对应的网络前缀地址=192.1.2.0),该 IP 分组被转发给由该路由项指定的下一跳路由器(路由器 R2),同样,路由器 R2 用 IP 地址 192.1.2.150 检索路由表时,判定该 IP 地址包含在由目的网络字段值 192.1.2.144/28 表示的 CIDR 地址块中(根据 28 位网络前缀计算出的目的 IP 地址 192.1.2.150 对应的网络前缀地址=192.1.2.144),该 IP 分组被转发给为第 6 组配置的网络。

图 5.12 中第 6 组对应的网络为了达到既提高访问外部网络的速度,又不改变自己的配置和访问其他组终端的速度的目的,采用同时连接路由器 R1 和路由器 R2 的方式,如图 5.13 所示。这种情况下,路由器 R1 中的路由项变为两项,分别指向路由器 R1 和第 6 组对应的网络。当路由器 R1 接收到目的 IP 地址为 192.1.2.150 的 IP 分组时,发现该 IP 地址与目的网络字段值 192.1.2.0/24 和 192.1.2.144/28 都匹配,路由器 R1 应该如何转发该 IP 分组?显然,路由器 R1 应该直接将该 IP 分组转发给第 6 组对应的网络,这也是将第 6 组对应的网络直接连接路由器 R1 的原因。路由器 R1 用最长前缀匹配来确定传输路径的优先级。最长前缀匹配是指如果有多个目的网络字段值和某个 IP 地址匹配,则选择网络前缀最长的目的网络字段值作为最终匹配结果。在路由器 R1 的路由项中目的网络字段值 192.1.2.0/24 的网络前缀是 24 位,而目的网络字段值 192.1.2.144/28 的网络前缀为 28 位,选择目的网络字段值 192.1.2.144/28 作为最终匹配结果。因此,将 IP 分组直接转发给第 6 组对应的网络。

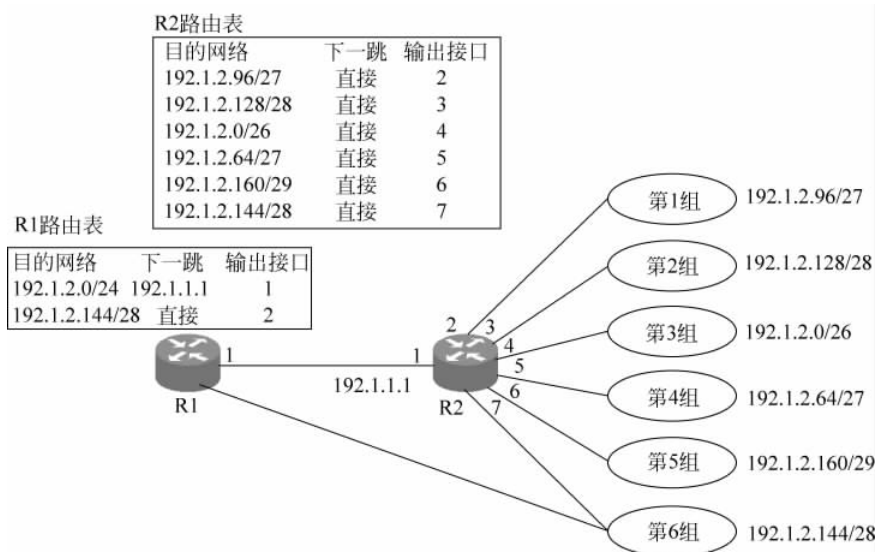


图 5.13 最长前缀匹配过程

5. 默认路由项

如果某个 IP 分组的目的 IP 地址和路由表中所有路由项的目的网络字段值均不匹配，或者丢弃该 IP 分组，或者选择默认路由项指定的传输路径。默认路由项的目的网络字段值为 0.0.0.0，对应的子网掩码为 0.0.0.0，表明所有 IP 地址都和默认网络地址 0.0.0.0/0.0.0.0 匹配。当通往多个网络的传输路径具有相同的下一跳时，可用一项默认路由项指明通往这些网络的传输路径。图 5.14 所示互连网络中，内部网络通过路由器 R1 连接 Internet，由于 Internet 由无数个网络组成，如果在路由器 R2 的路由表中详细列出 Internet 中所有网络对应的路由项，路由项数目将十分庞大。根据图 5.14 所示互连网络结构，路由器 R2 通往 Internet 的传输路径有着唯一的下一跳：路由器 R1，因此，除了用于指明通往内部网络的传输路径的路由项外，可用一项默认路由项指明通往 Internet 的传输路径。如果某个 IP 地址和 3 个内部网络的网络地址都不匹配，意味着该 IP 地址标识的目的终端位于 Internet，选择通往 Internet 的传输路径转发目的地址为该 IP 地址的 IP 分组。

6. 无分类编址与子网和超网

子网地址是无分类编址出现前的编址方式，一个单位需要分成若干组，每一个组出于安全考虑需要构建独立网络，但是每一个组的终端数较少，单独使用一个网络地址（如 C 类地址）会造成浪费，而且可能整个单位只分配到一个网络地址（如 C 类地址）。这种情况下，需要将单个网络地址分解为若干个子网地址。假如单位分配的 C 类地址是 192.1.1.0，需要将该 C 类地址均匀分配给 6 个子网，这样，每一个子网的主机号字段位数为 $8-3=5$ ，3 是子网号的位数，因为子网号的位数是满足 $2^n-2 \geq \text{子网数}$ 的最小 n ，之所以减 2，是因为当时规定子网字段值全 0 和全 1 都不能作为子网号。这种情况下，每一个子网对应的子网掩码为 255.255.255.224，6 个子网对应的子网地址分别是 192.1.1.32/255.255.255.224、

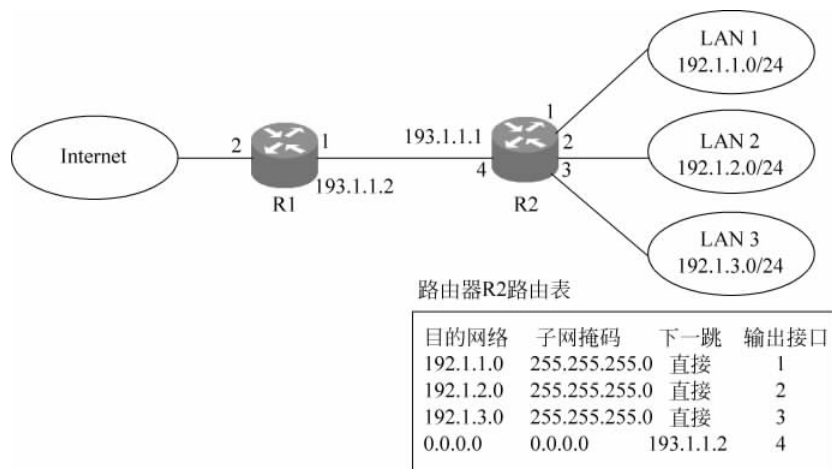


图 5.14 默认路由项功能

192.1.1.64/255.255.255.224、192.1.1.96/255.255.255.224、192.1.1.128/255.255.255.224、192.1.1.160/255.255.255.224、192.1.1.192/255.255.255.224。在出现无分类编址技术后,原来的地址类型已不复存在,32 位 IP 地址统一由网络前缀和主机号组成,不存在将某类网络地址划分为多个子网地址的问题。

出现无分类编址后,网络的主机号位数可以是 $N(N \geq 2)$,如果某个网络的终端数大于 $2^{10}-2$,小于等于 $2^{11}-2$,需要 11 位主机号,网络前缀位数为 $32-11=21$,可以选择网络地址 192.1.16.0/21。这个网络地址包含的 IP 地址范围恰好是 8 个 C 类地址 192.1.16.0~192.1.23.0 的 IP 地址的集合,将这样的网络地址称为超网地址。有时也将路由项聚合后生成的 CIDR 地址块称为超网地址,用于说明是多个网络的网络地址聚合后生成的 CIDR 地址块。无分类编址中任何由 IP 地址和网络前缀位数确定的 IP 地址集合的确切称呼是 CIDR 地址块,只是该 CIDR 地址块可以是单个网络的 IP 地址集合,也可以是多个网络的 IP 地址集合。

7. 举例

【例 5.3】 路由器的路由表如表 5.1 所示,回答以下问题:

① 假定路由器接收到目的 IP 地址为 142.150.71.132 的 IP 分组,给出路由器为该 IP 分组选择的下一跳,并说明理由。

② 在路由表中增加一项路由项,该路由项的作用是仅仅将目的 IP 地址为 142.150.71.132 的 IP 分组转发给下一跳: A,对目的 IP 地址为其他地址的 IP 分组的转发操作没有任何影响。

③ 在路由表中增加一项路由项,使所有目的 IP 地址中和路由表中已有路由项均不匹配的 IP 分组被转发给下一跳: E。

④ 将 142.150.64.0/24 划分为 4 个规模尽可能大的等长子网,给出每一个子网的可分配地址范围和子网掩码。

表 5.1 路由表

目的网络及前缀	下一跳
142.150.64.0/24	A
142.150.71.128/28	B
142.150.71.128/30	C
142.150.0.0/16	D

【解析】 ①根据路由项的目的网络字段值给出的网络前缀位数确定目的 IP 地址对应的网络前缀地址,目的 IP 地址 142.150.71.132 根据网络前缀位数 24、28、30 和 16 分别确定的网络前缀地址为 142.150.71.0、142.150.71.128、142.150.71.132 和 142.150.0.0。可以看出,目的 IP 地址 142.150.71.132 对应的网络前缀地址与路由项中目的网络字段值分别为 142.150.71.128/28 和 142.150.0.0/16 的路由项匹配,根据最长前缀匹配规则,选择路由项<142.150.71.128/28,B>作为最终匹配的路由项,确定下一跳为 B。这一问的关键是根据网络前缀长度求出目的 IP 地址对应的网络前缀地址,假定网络前缀长度为 N ,求出目的 IP 地址对应的网络前缀地址的过程如下,用最高 N 位为 1 的 32 位二进制数和 32 位目的 IP 地址进行“与”操作,得出的结果就是该目的 IP 地址对应的网络前缀地址。

② 要求路由项的目的网络字段值唯一匹配目的 IP 地址 142.150.71.132,确定增加的路由项为<142.150.71.132/32,A>。

③ 增加默认路由项<0.0.0.0/0,E>。

④ 把原来单个网络地址空间分成 4 个等长的网络地址空间,需要把原来的主机号位数减少 2 位,由 8 位变为 6 位,并用这减下的 2 位标识 4 个不同的子网,这样,每一个网络的网络号位数由 24 位变为 26 位,将 142.150.64.0/24 展开,用 26 位网络号重新划分子网后,可以得出如下 4 个子网的地址范围。

10001110 10010110 01000000 00 000000	}	子网 0,地址范围:142.150.64.0~142.150.64.63;
10001110 10010110 01000000 00 111111		可分配地址范围:142.150.64.1~142.150.64.62;
		子网掩码:255.255.255.192。
10001110 10010110 01000000 01 000000	}	子网 1,地址范围: 142.150.64.64~142.150.64.127;
10001110 10010110 01000000 01 111111		可分配地址范围: 142.150.64.65~142.150.64.126;
		子网掩码: 255.255.255.192。
10001110 10010110 01000000 10 000000	}	子网 2,地址范围: 142.150.64.128~142.150.64.191;
10001110 10010110 01000000 10 111111		可分配地址范围: 142.150.64.129~142.150.64.190;
		子网掩码: 255.255.255.192。
10001110 10010110 01000000 11 000000	}	子网 3,地址范围: 142.150.64.192~142.150.64.255;
10001110 10010110 01000000 11 111111		可分配地址范围: 142.150.64.193~142.150.64.254;
		子网掩码: 255.255.255.192。

5.2.4 IP 分组格式

1. 首部字段

IP 分组由首部与数据两部分组成。首部由 20 个字节的固定项和可变长度的可选项组

成。IP 分组首部格式如图 5.15 所示。



图 5.15 IP 分组首部格式

下面介绍 IP 分组首部各字段的意义。

版本：4b 版本字段给出 IP 分组所属 IP 协议的版本。由于每一个 IP 分组都含有版本字段，就允许存在几个月，甚至几年时间的升级过渡期，在这段时间内，不同版本的 IP 协议可同时在一个网络内运行。目前存在两种版本的 IP 协议：IPv4 和 IPv6，其版本号分别为 4 和 6。

首部长度：4b 首部长度字段以 32 位字为单位给出 IP 首部的实际长度。由于首部的长度不是固定的，需要用首部长度字段给出 IP 首部长度。字段最小值为 5，用于没有可选项的情况，由于 IP 首部长度的基本单位是 4 个字节，意味着首部固定部分长度为 20 字节。最大值为 15，这就将首部长度限制在 60 个字节内，意味着可选项长度不能超过 40 个字节。

服务类型：8b 服务类型字段允许终端告诉网络它希望得到的服务，可以通过服务类型字段指定 IP 分组的速率要求、可靠性要求及各种要求的组合。不同应用有不同的性能要求，对于数字化语音，快速到达比正确到达更重要。对于文件传送，无错传送比快速传送更为重要。服务类型字段从左到右包括三位优先级位，三位标志位 D、T、R 和目前没有使用的二位。三位优先级位表示从 0（普通报文）到 7（网络控制报文）8 级分组优先级，优先级高的 IP 分组优先得到服务。三位标志位允许终端指定最希望得到的服务。允许指定的服务是 D：时延，T：吞吐率，R：可靠性。D=1 表示该 IP 分组要求特别短的时延。T=1 表示该 IP 分组要求特别高的吞吐率。R=1 表示要求该 IP 分组尽可能不被损坏或丢弃。这些标志位可以帮助路由器选择对应的传输路径。实际上，早先的路由器一般都不考虑这些标志位，目前为支持多媒体应用，路由器开始支持服务分类（CoS）。

总长度：16b 总长度字段以字节为单位给出包括首部和数据的 IP 分组的长度，最大长度为 65535 字节。根据目前存在的传输网络状况，这个值是绰绰有余了。

标识：16b 标识字段告诉目的终端，那些数据片是属于同一 IP 分组的，属于同一 IP 分组的数据片具有相同标识字段值。发送端维持一个计数器，每发送一个 IP 分组，计数器加 1，计数器的值就作为 IP 分组的标识字段值。

标志：目前定义了 2 位标志位 DF 和 MF 位。DF 位置 1 要求不能对 IP 分组分片，它命令路由器不要把 IP 分组分片成数据片，因为目的终端没有能力把分片后的数据片重新装配成 IP 分组，例如：计算机引导时，ROM 要求把存储映像作为单个 IP 分组送给它。一旦 IP 分组中的 DF 位置 1，表明该 IP 分组只能作为单个数据片传送，这就要求路由器即使选择一条并不是最佳的路由，也要避开只能传输长度很短的 IP 分组的传输网络。要求所有网络至

少能传输小于 576 字节的 IP 分组。MF 位置 0 表示是若干数据片中最后一个数据片,除最后一个数据片外,IP 分组分片后所生成的所有其他数据片都必须将 MF 位设置为 1,MF 位的作用是使接收终端知道某个 IP 分组分片后所生成的所有数据片是否已全部接收到。

片偏移: 13b 片偏移字段以 8 个字节为单位给出该数据片在分片前的原始数据中的起始位置。因此,除最后数据片以外的所有其他数据片,它们的长度必须是 8 字节的整数倍。由于该字段有 13 位,由此可推出 IP 分组的最大长度为 $2^{13} \times 8 = 65536$ 字节。

生存时间: 字段长度 8b,此字段是用于限制 IP 分组存在时间的一个计数器,假定该计数器以秒为单位计数,IP 分组允许存在的最长时间为 255 秒。目前,该字段只是作为最大跳数使用,IP 分组每经过一跳路由器,该字段值减 1,当值减为 0 时,丢弃该 IP 分组并发送一个警告消息给源终端。设置该字段的目的是避免 IP 分组因为路由器的路由表被破坏而使 IP 分组在网络上无休止地漂荡。

协议: 字段长度 16b,当网络层把完整的 IP 分组装配好以后,它需要知道如何处理该 IP 分组。协议类型字段告诉网络层把该 IP 分组提交给哪一个进程处理。TCP 进程和 UDP 进程是最有可能处理该 IP 分组的进程。

首部检验和: 字段长度 16b,将 IP 分组首部以 16 位为单位分段,然后根据反码运算规则对各段进行累加,将累加结果求反作为首部检验和。首部检验和的作用是检测出首部传输过程中发生的错误。首部检验和经过每一跳路由器都必须重新计算一次,因为每经过一跳至少改变了一个首部字段值(生存时间字段)。

源地址和目的地址: 字段长度 32b,该字段给出了源终端的网络号和主机号及目的终端的网络号和主机号。

可选项: 设计该字段的目的是: ①允许以后协议版本提供原始设计中遗漏的信息; ②允许经验丰富的人试验一些新的想法; ③避免在报文首部中固定分配一些并不常用的信息字段。可选项长度可变。目前,定义了 5 种可选项,如表 5.2 所示,需要强调的是,并不是所有路由器都支持这 5 种可选项。

表 5.2 IP 可选项

可选项	描述
保密	指定 IP 分组如何保密
严格的源站选路	给出用于传输 IP 分组的完整路由
不严格的源站选路	给出不允许遗漏的一些路由器列表
记录路由	每一个经过的路由器将它的 IP 地址添加到 IP 分组中
时间戳	每一个经过的路由器将它的 IP 地址和时间戳添加到 IP 分组中

保密: 该选项给出如何保密 IP 分组,与军事应用有关的路由器可以用该选项来避开某些认为不安全的国家或地区。实际上,所有路由器都忽略该选项。

严格的源站选路: 该选项给出源终端至目的终端传输路径完整的 IP 地址列表,IP 分组必须严格遵循给出的传输路径。系统管理员可以用这种功能在路由器路由表损坏的情况下发送紧急 IP 分组,或者用于发送测量时间参数的 IP 分组。

不严格的源站选路: 该选项要求 IP 分组一定要经过列表中指定的路由器,并按指定的顺序经过。但允许通过传输路径上别的路由器。通过用该选项指定少数几个路由器来强迫

IP 分组经过某一特殊传输路径。例如：强迫从伦敦到悉尼的 IP 分组经过美国西部而不是东部时,该选项可指定 IP 分组必须经过纽约、洛杉矶、檀香山的路由器。当出于某种政治或经济考虑,需要 IP 分组经过或避开某些地区或国家时,可用该选项。

记录路由：该选项要求所有经过的路由器把它们的 IP 地址添加到该选项字段中,通过记录路由,可以帮助系统管理员查出路由算法中的一些问题。由于 ARPA 网刚建立时,IP 分组经过的路由器最多不超过 9 个,因此用 40 个字节记录经过的路由器已经很充足了,但对现在的 Internet 来说,用 40 个字节记录经过的路由器是远远不够的。

时间戳：该选项基本上与记录路由选项一样,不同的是,除记录 32 位的 IP 地址外,还记录 32 位的时间戳。该选项也主要用于诊断路由算法发生的错误。

IP 分组首部的可选项有很强地了解、管理网络的功能,常常被用来作为侦察网络的工具,为了网络的安全性,路由器需要关闭一些可选项的支持功能。

2. 分片

传输网络链路层帧净荷字段(也称载荷字段)允许的最大长度称为最大传输单元(Maximum Transfer Unit,MTU),如以太网的 MTU 为 1500B,如果 IP 分组长度超过传输该 IP 分组的传输网络的 MTU,必须将 IP 分组分片,分片过程是将 IP 分组净荷字段中的数据分片为多个数据片,除了最后一个数据片,其他数据片的长度必须是 8B 的整数倍。每一个数据片加上 IP 首部构成 IP 分组,必须保证分片后的数据片长度和 IP 首部长度之和小于传输网络的 MTU。通常情况下,除最后一个数据片,其他数据片长度的分配原则是：必须是 8 的倍数,且加上 IP 首部后尽量接近 MTU。为了标识这些由分片同一个 IP 分组净荷字段中的数据产生的 IP 分组序列,这些 IP 分组必须具有相同的标识字段值,为了在目的端将这些 IP 分组中净荷字段包含的数据片重新还原为原始数据,这些 IP 分组中的每一个 IP 分组必须在片偏移字段中给出该 IP 分组包含的数据片在原始数据中的起始位置,为了让目的端确定所有数据片对应的 IP 分组均已到达,必须标志最后一个数据片对应的 IP 分组。分片过程如图 5.16 所示,4000B 数据被分成 3 个数据片,长度分别是 1480B、1480B 和 1040B。

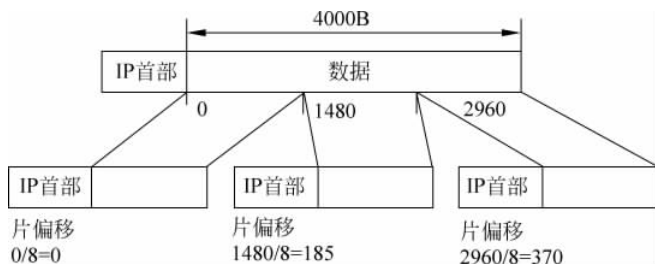


图 5.16 分片过程

【例 5.4】 终端 A 和终端 B 之间传输路径由网络 1、网络 2 和网络 3 组成,其中网络 1 的 MTU=1500B,网络 2 的 MTU=800B,网络 3 的 MTU=420B,假定终端 A 传输给终端 B 的数据的长度为 1440B,给出终端 A 及传输路径经过的各个路由器分片数据的过程。

【解析】 终端 A 及传输路径经过的路由器分片数据的过程如图 5.17 所示。终端 A 生成的 IP 分组的总长度为 1460B(包括 20B 首部和 1440B 净荷),由于终端 A 连接路由器 R1

的链路的 MTU=1500B,终端 A 可以直接将总长度为 1460B 的 IP 分组传输给路由器 R1。当路由器 R1 向路由器 R2 传输该 IP 分组时,发现输出链路的 MTU=800B,需要对 IP 分组进行分片操作。路由器 R1 将 IP 分组的净荷分成 2 个数据片,2 个数据片的长度分别为 776B 和 664B,加上 20B 的 IP 首部后,分别构成 2 个总长度分别为 796B(20B 首部+776B 净荷)和 684B 的 IP 分组。这 2 个 IP 分组的标识符字段值相同,后一个 IP 分组的片偏移=776/8=97。同样,当路由器 R2 向终端 B 传输这 2 个 IP 分组时,发现输出链路的 MTU=420B,路由器 R2 需要再一次对这 2 个 IP 分组进行分片操作,776B 的数据片被分片成长度分别为 400B 和 376B 的 2 个数据片,同样,664B 数据片被分片成长度分别为 400B 和 264B 的 2 个数据片,这 4 个数据片加上 IP 首部后构成 4 个 IP 分组,原来 M 标志位为 1 的 IP 分组分片后生成的 IP 分组序列的 M 标志位都为 1。原来 M 标志位为 0 的 IP 分组分片后生成的 IP 分组序列,除由最后一个数据片构成的 IP 分组外,其他 IP 分组的 M 标志位也都为 1。这些 IP 分组的标识符值都相同,图中每一个 IP 分组首部中的片偏移给出净荷中的数据片在原始净荷中的位置。

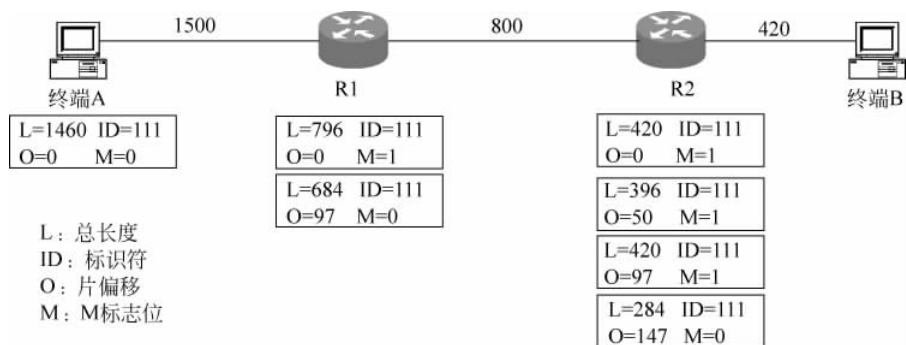


图 5.17 分片数据过程

5.3 路由表和 IP 分组端到端传输过程

5.3.1 路由表建立过程

1. 简单互连网络路由表生成过程

实现 IP 分组逐跳传输的关键是每一个路由器建立路由表,路由表中每一项路由项给出通往某个目的终端的传输路径。如果某个路由器通往一组目的终端的传输路径有着相同的下一跳,用一项路由项给出通往一组目的终端的传输路径。路由项的通用格式为<目的网络字段值,下一跳 IP 地址>,目的网络字段值给出表示一组目的终端的 IP 地址集合的网络前缀地址,下一跳地址值是下一跳连接互连当前跳和下一跳网络的接口的 IP 地址。

路由器每一个接口连接一个网络,该网络的网络地址由配置给该路由器接口的 IP 地址和子网掩码决定,如果某个路由器接口配置 IP 地址和子网掩码 192.1.1.254/255.255.0,意味着该接口连接的的网络的网络地址为 192.1.1.0/24。连接在该网络上的所有终

端的 IP 地址必须属于网络地址 192.1.1.0/24。对于图 5.18 所示网络结构,路由器 R1 接口 1 连接的网络 LAN 1 的网络地址为 192.1.1.0/24,路由器 R3 接口 1 连接的网络 LAN 2 的网络地址为 192.1.2.0/24,如果要求实现 LAN 1 中终端和 LAN 2 中终端之间的相互通信,必须建立 LAN 1 与 LAN 2 之间的双向传输路径。对于路由器 R1,LAN 1 是接口 1 直接连接的网络,通往 LAN 2 的传输路径上的下一跳是路由器 R2,路由器 R2 连接路由器 R1 的接口的 IP 地址是 192.2.1.2,因此,路由器 R1 分别生成用于指明通往网络 192.1.1.0/24 和网络 192.1.2.0/24 的传输路径的两项路由项 $\langle 192.1.1.0/24, \text{直接} \rangle$ 和 $\langle 192.1.2.0/24, 192.2.1.2 \rangle$ 。对于路由器 R2,通往 LAN 1 的传输路径上的下一跳是路由器 R1,路由器 R1 连接路由器 R2 的接口的 IP 地址是 192.2.1.1,通往 LAN 2 的传输路径上的下一跳是路由器 R3,路由器 R3 连接路由器 R2 的接口的 IP 地址是 192.2.2.2,因此,路由器 R2 分别生成用于指明通往网络 192.1.1.0/24 和网络 192.1.2.0/24 的传输路径的两项路由项 $\langle 192.1.1.0/24, 192.2.1.1 \rangle$ 和 $\langle 192.1.2.0/24, 192.2.2.2 \rangle$ 。对于路由器 R3,LAN 2 是接口 1 直接连接的网络,通往 LAN 1 的传输路径上的下一跳是路由器 R2,路由器 R2 连接路由器 R3 的接口的 IP 地址是 192.2.2.1,因此,路由器 R3 分别生成用于指明通往网络 192.1.1.0/24 和网络 192.1.2.0/24 的传输路径的两项路由项 $\langle 192.1.1.0/24, 192.2.2.1 \rangle$ 和 $\langle 192.1.2.0/24, \text{直接} \rangle$ 。各个路由器最终生成的用于指明通往网络 LAN 1 和 LAN 2 的传输路径的路由项如图 5.18 所示。

路由器通过路由表确定通往网络 LAN 1 和 LAN 2 的传输路径,连接 LAN 1 上的终端通过默认网关地址确定通往 LAN 2 的传输路径,该默认网关地址是路由器 R1 连接 LAN 1 的接口(接口 1)的 IP 地址。终端的默认网关地址给出该终端通往其他网络的传输路径上的第一跳路由器地址,该终端传输给其他网络上终端的 IP 分组,首先传输给与其连接在同一个网络上的、用默认网关地址指定的第一跳路由器。如果有多个路由器连接在终端所连接的网络上,可以选择其中一个路由器作为第一跳路由器,用该路由器连接终端所在网络的接口的 IP 地址作为终端的默认网关地址。

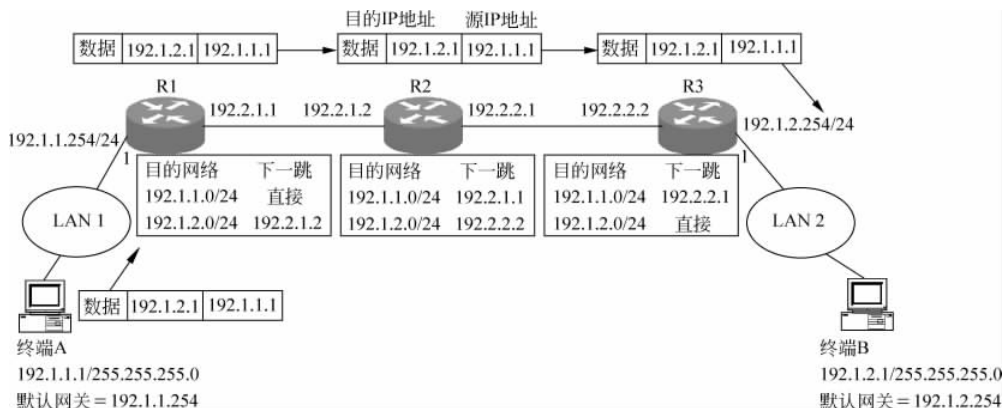


图 5.18 IP 分组传输过程

2. 复杂互连网络路由表生成过程

复杂互连网络结构如图 5.19 所示。假定路由器 R5 要为 4 个网络(网络地址分别为

192.1.1.0/24, 192.1.2.0/24, 192.1.3.0/24, 192.1.4.0/24) 选择最少跳数的传输路径, 即经过的路由器数目最少的传输路径(这种传输路径也称为最短路径)。管理员可以通过分析图 5.19 所示的互连网络拓扑结构, 确定路由器 R5 通往这 4 个网络的最短路径, 如图 5.19 中箭头指出的传输路径, 并因此给出表 5.3 所示的路由表, 其中目的网络字段给出需要到达的网络的网络地址, 距离字段给出路由器 R5 到达目的网络的最短路径所经过的路由器数目(含路由器 R5), 下一跳字段给出路由器 R5 通往目的网络的最短路径上的下一跳路由器, 如果目的网络和路由器 R5 直接相连, 下一跳字段中用直接表示。

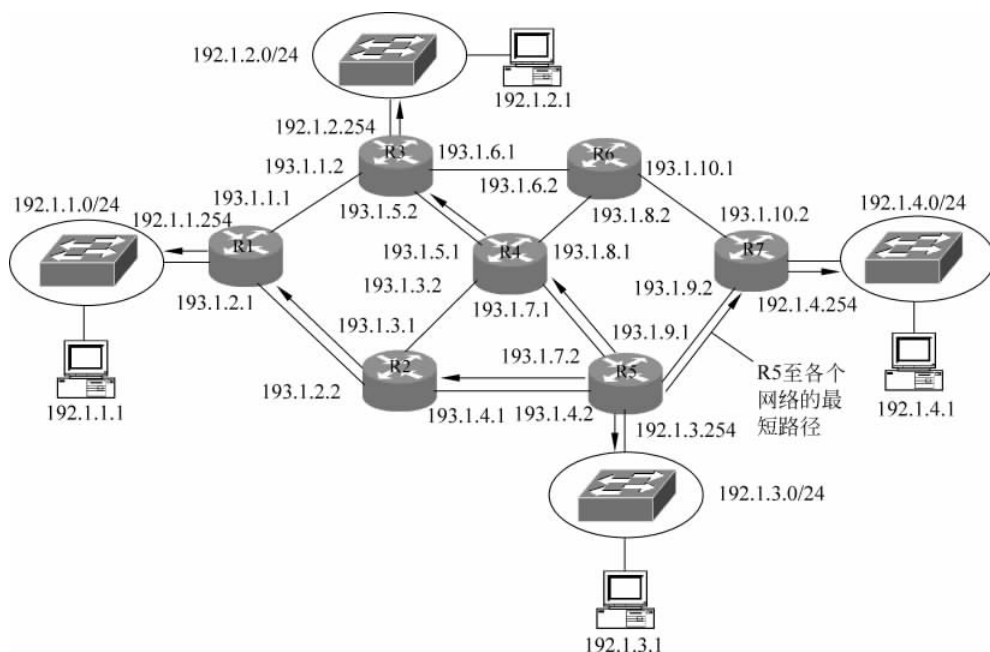


图 5.19 人工配置静态路由表过程

表 5.3 路由器 R5 配置的静态路由表

目的网络	距离	下一跳路由器
192.1.1.0/24	3	193.1.4.1
192.1.2.0/24	3	193.1.7.1
192.1.3.0/24	1	直接
192.1.4.0/24	2	193.1.9.2

3. 例题解析

【例 5.5】 互连网络结构如图 5.20 所示, 路由表每一项路由项包含字段<目的网络, 子网掩码, 下一跳, 输出接口>, 回答下列问题:

① 将 IP 地址空间 202.115.1.0/24 划分为两个子网, 分别分配给 LAN 1 和 LAN 2, 每个子网分配的 IP 地址数不少于 120, 给出子网划分结果, 说明理由并给出必要的计算过程;

② 给出路由器 R1 的路由表, 包含用于指明通往图 5.20 中所有网络和服务器的传输路径的路由项;

③ 给出路由器 R2 路由表中用于指明通往 LAN 1 和 LAN 2 的传输路径的路由项。

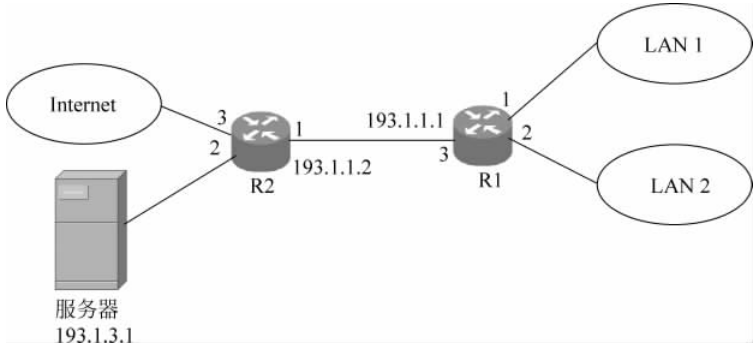


图 5.20 互连网络结构

【解析】 ① 由于要求每一个子网的 IP 地址数不小于 120,因此,每一个子网至少用 7 位二进制数作为主机号,这样,可以把给定的 IP 地址空间分成两个等长的网络地址空间,把原来的主机号位数减少 1 位,由 8 位变为 7 位,并用这减下的 1 位标识两个不同的子网,这样,每一个网络地址空间中的网络号位数由 24 位变为 25 位,将 202.115.1.0/24 展开,用 25 位网络号重新划分子网后,可以得出如下两个子网的 IP 地址范围。

子网 0,地址范围: 202.115.1.0~

11001010 01110011 00000001 0 0000000

202.115.1.127;

10001110 10010110 01000000 0 1111111

网络地址: 202.115.1.0;

子网掩码: 255.255.255.128。

子网 1,地址范围:202.115.1.128~

11001010 01110011 00000001 1 0000000

202.115.1.255;

10001110 10010110 01000000 1 1111111

网络地址: 202.115.1.128;

子网掩码: 255.255.255.128。

② 路由器 R1 的路由表如表 5.4 所示。由于两个子网和路由器 R1 直接相连,表中用于指明通往两个子网的传输路径的路由项的下一跳字段值为直接。用于指明通往服务器的传输路径的路由项中的目的网络和子网掩码字段值必须唯一匹配该服务器的 IP 地址。因此,用 32 位全 1 的子网掩码,表示目的网络只包含单个 IP 地址: 193.1.3.1。由于 Internet 是由无数个网络组成的,因此,只能以默认路由项指明通往 Internet 的传输路径。由于路由器 R1 至服务器和 Internet 传输路径上的下一跳是路由器 R2,且路由器 R2 接口 1 和路由器 R1 接口 3 连接在同一个网络上,用路由器 R2 接口 1 的 IP 地址作为这些路由项的下一跳 IP 地址。

表 5.4 路由器 R1 路由表

目的网络	子网掩码	下一跳	输出接口
202.115.1.0	255.255.255.128	直接	1
202.115.1.128	255.255.255.128	直接	2
193.1.3.1	255.255.255.255	193.1.1.2	3
0.0.0.0	0.0.0.0	193.1.1.2	3

③ 路由器 R2 应该有两项路由项分别用于指明通往 LAN 1 和 LAN 2 的传输路径,但这两项路由项有着相同的输出接口和下一跳,这样的路由项可以尝试聚合为一项路由项,前提是聚合后的目的网络和子网掩码字段确定的 IP 地址空间等于聚合前两项路由项所包含的 IP 地址空间,这里, $(202.115.1.0 \sim 202.115.1.127) + (202.115.1.128 \sim 202.115.1.255) = (202.115.1.0 \sim 202.115.1.255)$,因此,可以完成图 5.21 所示的合并过程。

目的网络	子网掩码	下一跳	接口	目的网络	子网掩码	下一跳	接口
202.115.1.0	255.255.255.128	193.1.1.1	1	202.115.1.0	255.255.255.0	193.1.1.1	1
202.115.1.128	255.255.255.128	193.1.1.1	1				

图 5.21 路由项合并过程

5.3.2 IP 分组端到端传输过程

为了深刻理解路由表的作用,详细讨论图 5.18 中终端 A 至终端 B 的 IP 分组传输过程。

1. 确定源终端和目的终端是否在同一个网络

终端 A 向终端 B 传输数据前,必须先获取终端 B 的 IP 地址,然后将数据封装成以终端 A 的 IP 地址为源 IP 地址,以终端 B 的 IP 地址为目的 IP 地址的 IP 分组,在进行 IP 分组传输前,先确定终端 B 是否和终端 A 位于同一个网络,步骤如下:

- (1) 终端 A 根据自己的 IP 地址和子网掩码,求出网络地址: 192.1.1.0。
- (2) 终端 A 根据终端 B 的 IP 地址和自己的子网掩码,求出终端 B 的网络地址: 192.1.2.0。
- (3) 如果两个网络地址相同,说明终端 A 和终端 B 位于同一个网络,终端 A 至终端 B 的 IP 分组传输过程无须经过路由器。
- (4) 如果两个网络地址不相同,说明终端 A 和终端 B 位于不同的网络,终端 A 将沿着由路由器构成的 IP 分组传输路径,逐跳转发 IP 分组。

2. 根据默认网关地址找到第一跳路由器

一旦确定终端 B 和终端 A 不在同一个网络,终端 A 将 IP 分组转发给终端 A 至终端 B 传输路径上的第一跳路由器,该路由器的 IP 地址通过配置的默认网关地址获得,这里的默认网关实际上是默认路由器,因此,也将默认网关地址称为默认路由器地址。如果连接终端 A 和第一跳路由器的网络是以太网,必须将 IP 分组封装成以终端 A 的 MAC 地址为源 MAC 地址,以第一跳路由器连接以太网的接口的 MAC 地址为目的 MAC 地址的 MAC 帧,然后将 MAC 帧通过以太网传输给第一跳路由器。

3. 路由器逐跳转发

IP 分组到达路由器 R1 后,路由器 R1 根据最长前缀匹配算法用 IP 分组的目的 IP 地址匹配路由表中的所有路由项,如果找到匹配的路由项,将 IP 分组转发给路由项指定的下一跳。路由器 R1 的路由表中,只有路由项 $\langle 192.1.2.0/24, 192.2.1.2 \rangle$ 和 IP 分组的目的 IP

地址匹配,IP 分组被转发给 IP 地址为 192.2.1.2 的下一跳路由器。传输路径上的路由器依次逐跳转发,IP 分组到达传输路径上最后一跳路由器 R3。

4. 直接交付

路由器 R3 中和 IP 分组目的 IP 地址匹配的路由项是<192.1.2.0/24,直接>,表明该路由器和终端 B 之间不再有其他路由器,即终端 B 和该路由器的其中一个接口连接在同一个网络上,路由器通过该网络将 IP 分组直接传输给终端 B。如果连接路由器 R3 和终端 B 的网络是以太网,必须将 IP 分组封装成以路由器 R3 连接以太网的接口的 MAC 地址为源 MAC 地址、以终端 B 的 MAC 地址为目的 MAC 地址的 MAC 帧,然后将 MAC 帧通过以太网传输给终端 B。

从上述讨论的 IP 分组端到端传输过程可以得出以下实现 IP 分组端到端传输的基本思路:

(1) 建立一条以源终端为始点,以目的终端为终点,中间由若干路由器组成的 IP 分组端到端传输路径,IP 分组沿着端到端传输路径逐跳转发。源终端通过配置的默认网关地址获得第一跳路由器的 IP 地址,中间路由器根据路由表和 IP 分组的目的 IP 地址确定下一跳路由器地址;

(2) 在获取下一跳路由器的 IP 地址后,通过 IP over X 技术,实现 IP 分组当前跳至下一跳的传输过程,X 是连接当前跳和下一跳的传输网络,如以太网。

建立端到端传输路径的关键是每一个路由器建立路由表,路由表中每一项路由项指出通往特定网络的传输路径上的下一跳路由器,因此,解决 IP 分组端到端传输的第一步是为互连网络中的每一个路由器建立路由表。

5. 例题解析

【例 5.6】 确定下述主机对是否连接在同一个网络上。

- ① 主机 1: 172.16.5.72/255.255.255.0, 主机 2: 172.16.5.79/255.255.255.0。
- ② 主机 1: 192.168.19.35/255.255.255.224, 主机 2: 192.168.19.48/255.255.255.224。
- ③ 主机 1: 10.128.14.14/255.255.255.240, 主机 2: 10.128.14.19/255.255.255.240。
- ④ 主机 1: 192.168.3.68/255.255.255.248, 主机 2: 192.168.3.74/255.255.255.248。

【解析】 判定两个主机是否连接在同一个网络上的依据是这两个主机的网络地址是否相同,主机的网络地址是主机的 IP 地址与主机的子网掩码进行“与”操作的结果。

① 由于子网掩码是 24 位 1 和 8 位 0,主机的网络地址取 IP 地址的前 24 位,最后 8 位为 0,因此,主机 1 的网络地址=172.16.5.0/24,主机 2 的网络地址=172.16.5.0/24,主机 1 和主机 2 连接在同一个网络上。

② 由于子网掩码是 27 位 1 和 5 位 0,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 3 位值。

$$00100011(35) \& \& 11100000(224) = 00100000(32)$$

$$00110000(48) \& \& 11100000(224) = 00100000(32)$$

由此得出主机 1 和主机 2 的网络地址均是 192.168.19.32/27。主机 1 和主机 2 连接在同一个网络上。

③ 由于子网掩码是 28 位 1 和 4 位 0,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 4 位值。

$$00001110(14) \& . \& . 11110000(240) = 00000000(0)$$

$$00010011(19) \& . \& . 11110000(240) = 00010000(16)$$

由此得出主机 1 的网络地址 = 10.128.14.0/28,主机 2 的网络地址 = 10.128.14.16/28。主机 1 和主机 2 连接在不同的网络上。

④ 由于子网掩码是 29 位 1 和 3 位 0,且两个 IP 地址的前 24 位相同,因此,只需计算出 IP 地址最后 1 个字节的前 5 位值。

$$01000100(68) \& . \& . 11111000(248) = 01000000(64)$$

$$01001010(74) \& . \& . 11111000(248) = 01001000(72)$$

由此得出主机 1 的网络地址 = 192.168.3.64/29,主机 2 的网络地址 = 192.168.3.72/29。主机 1 和主机 2 连接在不同的网络上。

5.3.3 ARP 和地址解析过程

1. 地址解析过程

图 5.22 中,假定终端 A 和服务器 B 连接在同一个网络上,即使如此,终端 A 访问服务器 B 时所给出的也不会是服务器 B 的 MAC 地址,往往是服务器 B 的域名,经过域名服务器解析后得到的也只能是服务器 B 的 IP 地址。根据以太网交换机的工作原理,以太网交换机只能根据 MAC 帧的目的 MAC 地址和转发表来转发 MAC 帧,这就意味着:①不能在以太网上直接传输 IP 分组,必须将 IP 分组封装成 MAC 帧;②在将 IP 分组封装成 MAC 帧前,必须先获取连接在同一个网络上的源终端和目的终端的 MAC 地址。源终端的 MAC 地址可以直接从安装的网卡中读取,问题是如何根据目的终端的 IP 地址来获取目的终端的 MAC 地址。地址解析协议(Address Resolution Protocol,ARP)和地址解析过程就用于实现这一功能,ARP 请求帧格式如图 5.23 所示。

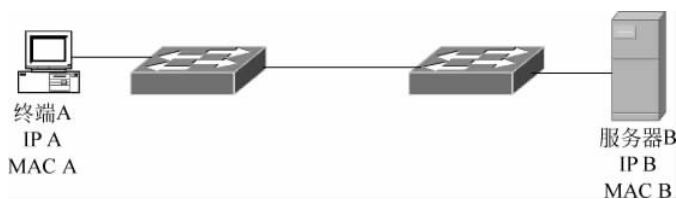


图 5.22 以太网内传送 IP 分组过程

图 5.24 中,终端 A 获知了服务器 B 的 IP 地址 IP B 后,广播一个 MAC 帧,该 MAC 帧的格式如图 5.23 所示,它的源 MAC 地址为终端 A 的 MAC 地址 MAC A,目的 MAC 地址为广播地址 ff-ff-ff-ff-ff-ff,MAC 帧中的数据字段包含终端 A 的 IP 地址 IP A 和 MAC 地址 MAC A,同时,包含服务器 B 的 IP 地址 IP B,IP B 是需要解析的 IP 地址,称为目标地址。该帧是 ARP 请求帧,它要求 IP 地址为 IP B 的网络终端回复它的 MAC 地址。

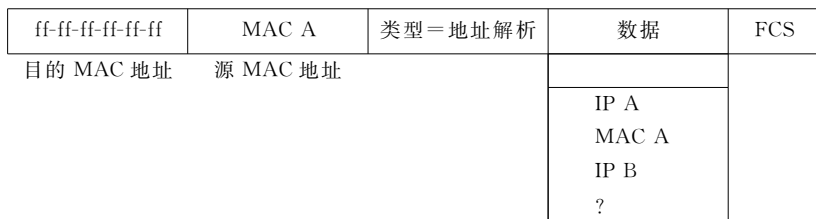


图 5.23 用于地址解析的 MAC 帧

由于该 MAC 帧的目的地址为广播地址,同一网络内的所有终端都能够接收到该 MAC 帧,每一个接收到该 MAC 帧的终端首先检测自己的 ARP 缓冲区,如果 ARP 缓冲区中没有发送终端的 IP 地址和 MAC 地址对,将发送终端的 IP 地址和 MAC 地址对(IP A 和 MAC A)记录在 ARP 缓冲区中,然后比较 MAC 帧中给出的目标 IP 地址是否和自己的 IP 地址相同,如果相同,回复自己的 MAC 地址,整个过程如图 5.24 所示。

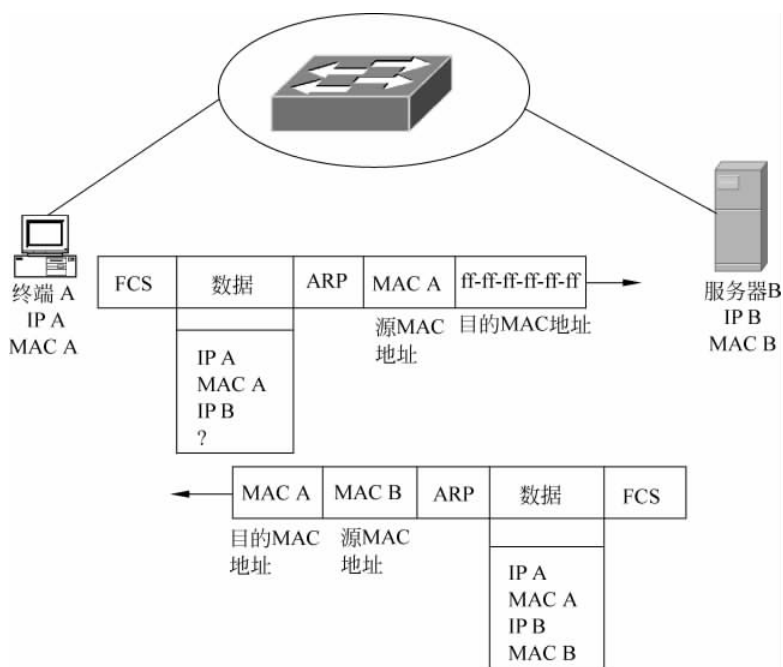


图 5.24 ARP 解析地址过程

ARP 地址解析过程只能发生在连接在同一个以太网上的源终端和目的终端之间,如果源终端和目的终端不在同一个网络内,则 IP 分组需要逐跳转发,源终端必须先将 IP 分组发送给由默认网关地址指定的第一跳路由器,当然,如果连接源终端和第一跳路由器的网络是以太网,源终端通过 ARP 地址解析过程获取第一跳路由器连接以太网的接口的 MAC 地址。同样,如果连接第一跳和下一跳路由器的网络也是以太网,如图 5.25 所示,第一跳路由器也需通过 ARP 地址解析过程获取下一跳路由器连接以太网的接口的 MAC 地址。总之,如果互连当前跳和下一跳的网络是以太网,IP 分组封装成 MAC 帧后才能经过以太网实现当前跳至下一跳的传输过程,在将 IP 分组封装成 MAC 帧前,必须获取下一跳连接以太网

的接口的 MAC 地址,ARP 地址解析过程用于完成根据下一跳连接以太网的接口的 IP 地址求出该接口的 MAC 地址的过程。

2. MAC 帧封装过程

图 5.25 给出了终端 A 传输给终端 B 的 IP 分组经过各个以太网时的封装过程,IP 分组由终端 A 至终端 B 的传输过程中是不变的,但 IP 分组经过互连终端 A 和路由器 R1 的以太网时,封装成以路由器 R1 接口 1 的 MAC 地址 MAC R11 为目的 MAC 地址、以终端 A 的 MAC 地址 MAC A 为源 MAC 地址的 MAC 帧,类型字段 0800 表示净荷是 IP 分组。终端 A 通过解析默认网关地址 192.1.1.254 获得路由器 R1 接口 1 的 MAC 地址。IP 分组经过互连路由器 R1 和路由器 R2 的以太网时,封装成以路由器 R2 接口 1 的 MAC 地址 MAC R21 为目的 MAC 地址、以路由器 R1 接口 2 的 MAC 地址 MAC R12 为源 MAC 地址的 MAC 帧。路由器 R1 通过检索路由表获取路由器 R2 接口 1 的 IP 地址 192.1.3.2,通过解析 IP 地址 192.1.3.2 获得路由器 R2 接口 1 的 MAC 地址。IP 分组经过互连路由器 R2 和终端 B 的以太网时,封装成以终端 B 的 MAC 地址 MAC B 为目的 MAC 地址、以路由器 R2 接口 2 的 MAC 地址 MAC R22 为源 MAC 地址的 MAC 帧。路由器 R2 通过检索路由表得知终端 B 直接连接在接口 2 连接的以太网上,通过解析终端 B 的 IP 地址 192.1.2.1 获得终端 B 的 MAC 地址。

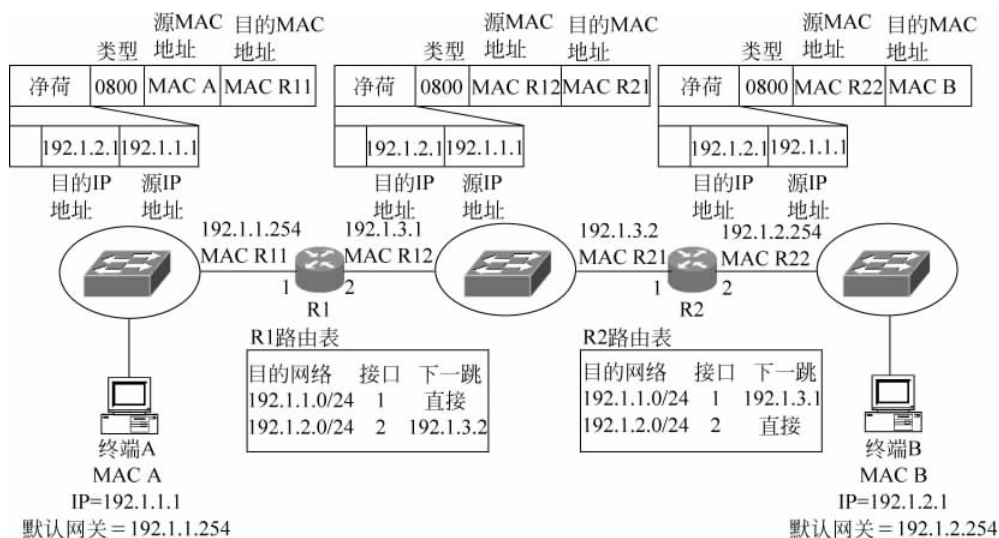


图 5.25 由多个以太网互连而成的互联网

5.4 虚拟路由器冗余协议

5.4.1 容错网络结构

互连网络结构如图 5.26 所示,每一个以太网内部通过链路冗余和生成树协议保证在发生单条链路故障的情况下仍然保持连接在同一以太网上的终端之间的连通性。同时,路由

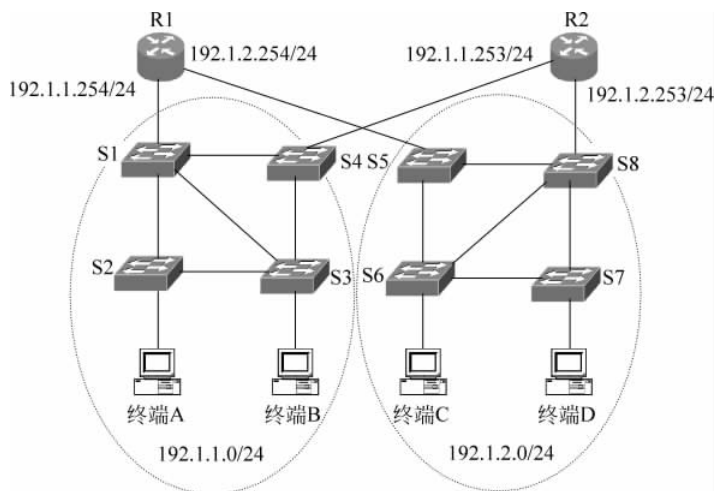


图 5.26 容错网络结构

器 R1 和路由器 R2 分别有接口连接到两个以太网,保证在其中一个路由器发生故障的情况下仍然保持连接在不同以太网上的终端之间的连通性,因此,图 5.26 所示互连网络结构是一种不会因为单点故障导致网络连通性发生问题的容错网络结构。

由于每一个以太网同时连接两个路由器接口,因此,连接在每一个以太网上的终端可以在分配给两个路由器接口的两个 IP 地址中选择一个 IP 地址作为默认网关地址。如终端 A 可以选择 192.1.1.254 或 192.1.1.253 作为默认网关地址,但由于目前终端一般只能配置一个默认网关地址,因此,即使对于图 5.26 所示容错结构,终端在只能配置单个默认网关地址且作为默认网关的路由器失效的情况下,必须通过手工配置新的默认网关地址来保持该终端和其他终端之间的连通性,如果终端 A 配置了默认网关地址 192.1.1.254,一旦路由器 R1 失效,必须通过手工配置方式为终端 A 配置新的默认网关地址 192.1.1.253,否则,终端 A 无法和连接在其他网络上的终端通信。

对于图 5.26 所示的容错网络结构,希望有一种和生成树协议相似的协议,该协议能够根据优先级在多个可以作为默认网关的路由器中选择一个路由器作为其默认网关,一旦该路由器发生故障,能够自动选择另一个路由器作为默认网关,并自动完成两个路由器之间的功能切换。虚拟路由器冗余协议(Virtual Router Redundancy Protocol,VRRP)就是这样一种协议。

5.4.2 VRRP 工作原理

1. VRRP 工作环境

VRRP 工作环境如图 5.27 所示,支持 VRRP 的路由器称为 VRRP 路由器,多个有接口连接在同一个网络上的 VRRP 路由器(如图 5.27 中路由器 R1 和路由器 R2)构成一个虚拟路由器,这些 VRRP 路由器中只有一个 VRRP 路由器是主路由器,其他路由器为备份路由器。VRRP 作用的网络可以是任意支持广播的网络,如以太网、令牌环网和 FDDI,连接在这些网络上的终端和路由器接口有着唯一的 MAC 地址,这里以以太网为例来讨论 VRRP 的

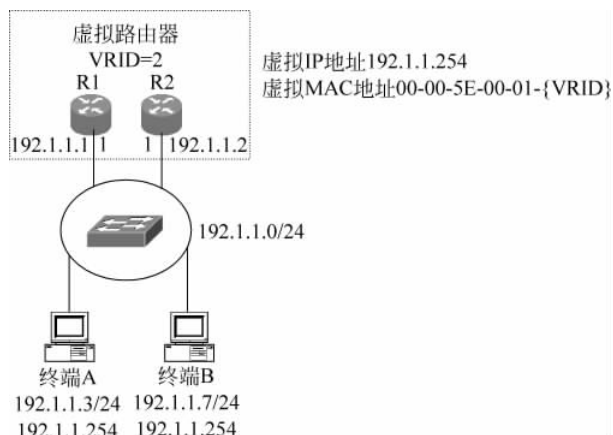


图 5.27 VRRP 工作环境

工作原理。每一个 VRRP 路由器连接以太网的接口可以分配多个 IP 地址,从这些 IP 地址中选择一个作为接口的基本 IP 地址,接口发送的 VRRP 报文以接口的基本 IP 地址作为 IP 分组的源 IP 地址。可以对虚拟路由器配置多个 IP 地址,这些 IP 地址称为虚拟 IP 地址,虚拟 IP 地址可以与为 VRRP 路由器接口配置的 IP 地址相同,如果某个 VRRP 路由器为某个接口配置的 IP 地址与为该接口所属的虚拟路由器配置的虚拟 IP 地址相同,该路由器称为 IP 地址拥有者。每一个虚拟路由器分配唯一的 8 位二进制数的虚拟路由器标识符(Virtual Router Identifier,VRID),属于同一个虚拟路由器的多个 VRRP 路由器有着相同的虚拟路由器标识符。虚拟路由器对外有着唯一的 MAC 地址 00-00-5E-00-01-{VRID},对于 VRID 为 2 的虚拟路由器,虚拟 MAC 地址为 00-00-5E-00-01-02。终端配置的默认网关地址必须是虚拟 IP 地址,对虚拟 IP 地址进行地址解析得到的结果必须是虚拟 MAC 地址,以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧一定能够到达主路由器,只有主路由器转发封装在以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧中的 IP 分组。

VRRP 需要解决的问题主要有以下三项。

- 在属于同一个虚拟路由器的多个 VRRP 路由器中产生主路由器;
- 一旦接收到终端发送的请求解析虚拟 IP 地址的 ARP 请求报文,虚拟路由器将虚拟 MAC 地址作为与虚拟 IP 地址绑定的 MAC 地址回送给终端;
- 以太网(严格地讲是所有支持广播的局域网)一定能够将以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧送达主路由器。

2. 路由器初始配置

对于图 5.27 所示的 VRRP 工作环境,路由器 R1 和路由器 R2 需要完成以下基本配置。

- 分别在路由器 R1 和路由器 R2 创建 VRID 为 2 的虚拟路由器,分别将路由器 R1 和路由器 R2 的接口 1 配置给 VRID 为 2 的虚拟路由器,使得路由器 R1 和路由器 R2 成为 VRID 为 2 的虚拟路由器的 VRRP 路由器;
- 分别为路由器 R1 和路由器 R2 的接口 1 分配 IP 地址 192.1.1.1/24 和 192.1.1.2/24,这两个接口的 IP 地址必须与它们连接的以太网的网络地址 192.1.1.0/24 一

致,由于路由器 R1 和路由器 R2 的接口 1 只分配了一个 IP 地址,该 IP 地址称为接口的基本 IP 地址;

- 为路由器 R1 和路由器 R2 的接口 1 分配优先级,优先级的范围为 1~254,主路由器用优先级 0 表示愿意主动放弃主路由器地位,IP 地址拥有者的优先级为 255。优先级值高的 VRRP 路由器在竞争主路由器时具有较高优先级;
- 为 VRID 为 2 的虚拟路由器分配虚拟 IP 地址 192.1.1.254。该 IP 地址成为连接在网络 192.1.1.0/24 上的终端的默认网关地址;
- 虚拟路由器根据 VRID=2 生成虚拟 MAC 地址 00-00-5E-00-01-02。

3. VRRP 报文格式

VRRP 报文封装成 IP 分组的格式如图 5.28 所示,不直接将 VRRP 报文封装成 MAC 帧格式的主要原因是 VRRP 作用的网络可以是支持广播的任意网络,不一定是以太网。IP 分组的源 IP 地址是发送 VRRP 报文的接口的基本 IP 地址,对于路由器 R1 接口 1 发送的 VRRP 报文,其源 IP 地址为 192.1.1.1,目的 IP 地址是组播地址 224.0.0.18。所有 VRRP 路由器将以该组播地址为目的地址的 IP 分组提交给 VRRP 实体。VRRP 报文对应的协议字段值是 112。VRRP 报文中给出发送该 VRRP 报文的接口所属的虚拟路由器的 VRID、该接口的优先级、分配给虚拟路由器的虚拟 IP 地址等。VRRP 只有一种类型报文——通告报文。

源 IP 地址	目的 IP 地址	协议	
192.1.1.1	224.0.0.18	112	净荷
			VRID=2
			优先级
			虚拟 IP 地址 (192.1.1.254)

图 5.28 VRRP 报文格式

如果 VRRP 作用的网络是以太网,图 5.28 所示的 IP 分组将封装成 MAC 帧,该 MAC 帧的源 MAC 地址是发送接口所属虚拟路由器对应的虚拟 MAC 地址,对于路由器 R1 接口 1,源 MAC 地址是 00-00-5E-00-01-02,目的 MAC 地址是组播地址 224.0.0.18 对应的 MAC 组地址。根据组播地址 224.0.0.18 求出对应的 MAC 组地址的过程如图 5.29 所示。

高5位		低23位	
1110			
00000001	00000000	01011110	0

图 5.29 IP 组播地址映射到 MAC 组地址过程

从图 5.29 中可以看出,映射后的 MAC 地址的高 25 位固定为 00000001、00000000、01011110 和 0,低 23 位等于组播地址的低 23 位。因此,组播地址 224.0.0.18 对应的 MAC 组地址为 01-00-5E-00-00-12。由于组播地址中用于标识组播组的地址有 28 位,因此,标识组播组的组播地址中的高 5 位在映射过程中没有使用,这就使得组播地址和 MAC 组地址之间的映射不是唯一的,32 个不同的组播地址有可能映射为同一个 MAC 组地址。

4. 主路由器产生过程

路由器状态转换过程如图 5.30 所示,每一个 VRRP 路由器启动后,处于初始化状态,如果该 VRRP 路由器是 IP 地址拥有者,该 VRRP 路由器立即成为主路由器,并立即发送图 5.28 所示的 VRRP 报文,然后,周期性地发送 VRRP 报文。如果某个 VRRP 路由器不是 IP 地址拥有者,该 VRRP 路由器立即成为备份路由器,启动 Master_Down_Timer,等待接收主路由器发送的 VRRP 报文。

任何路由器接收到 VRRP 报文后,依序进行下列检查。

- 判别接收该 VRRP 报文的接口是否属于 VRRP 报文中 VRID 指定的虚拟路由器;
- 根据 VRRP 报文中的 VRID 确定虚拟路由器,判别路由器为该虚拟路由器配置的虚拟 IP 地址是否与 VRRP 报文中给出的虚拟 IP 地址相同。

上述检查中只要有一项不匹配,路由器将丢弃该 VRRP 报文。

如果主路由器接收到 VRRP 报文,而且 VRRP 报文中的优先级大于主路由器为接收该 VRRP 报文的接口配置的优先级,或者虽然 VRRP 报文中的优先级等于主路由器为接收该 VRRP 报文的接口配置的优先级,但 VRRP 报文的源 IP 地址大于主路由器接收该 VRRP 报文的接口的基本 IP 地址,该主路由器立即转换为备份路由器,停止发送 VRRP 报文,启动 Master_Down_Timer,等待新的主路由器发送 VRRP 报文。

备份路由器接收到主路由器发送的 VRRP 报文后,根据备份路由器的工作方式对 VRRP 报文进行处理,如果备份路由器配置为允许抢占方式,且发现 VRRP 报文中的优先级小于备份路由器为接收该 VRRP 报文的接口配置的优先级,备份路由器立即转换为主路由器,并立即发送 VRRP 报文,然后,周期性地发送 VRRP 报文。如果备份路由器配置为不允许抢占方式,或者发现 VRRP 报文中的优先级大于或等于备份路由器为接收该 VRRP 报文的接口配置的优先级,刷新 Master_Down_Timer。

如果某个备份路由器的 Master_Down_Timer 溢出,表示主路由器已经失效,该备份路由器立即转换为主路由器,并立即发送 VRRP 报文,然后,周期性地发送 VRRP 报文。有可能因为网络拥塞导致主路由器发送的 VRRP 报文不能及时到达备份路由器,因而使备份路由器误认为主路由器失效而重新开始主路由器选择过程,为了避免发生这种情况,Master_Down_Timer 溢出时间大于 $3 \times$ 主路由器 VRRP 报文发送间隔。

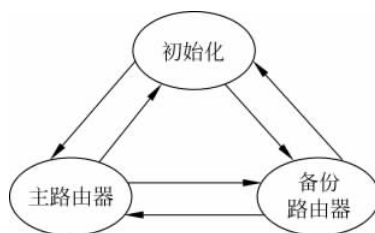


图 5.30 路由器状态转换过程

5. 主路由器和备份路由器功能

1) 主路由器功能

- 必须对请求解析虚拟 IP 地址的 ARP 请求报文做出响应；
- 必须对封装在以虚拟 MAC 地址为目的 MAC 地址的 MAC 帧中的 IP 分组进行转发操作；
- 在成为主路由器时,立即发送将所有虚拟 IP 地址绑定到虚拟 MAC 地址的 ARP 报文,使得网络内的所有终端将默认网关地址与虚拟 MAC 地址绑定在一起。

2) 备份路由器功能

- 不对请求解析虚拟 IP 地址的 ARP 请求报文做出响应；
- 丢弃接收到的以虚拟 MAC 地址为目的地址的 MAC 帧；
- 丢弃接收到的以虚拟 IP 地址为目的地址的 IP 分组。

6. 虚拟 IP 地址解析过程

如果终端在 ARP 缓存中找不到与默认网关地址绑定的 MAC 地址,会发送一个请求解析该默认网关地址的 ARP 请求报文,该 ARP 请求报文在终端所连接的网络中广播,连接在该网络上的所有 VRRP 路由器都接收到该 ARP 请求报文,但只有主路由器对该 ARP 请求报文做出响应,并在 ARP 响应报文中将虚拟 MAC 地址与默认网关地址绑定在一起。终端发送给默认网关的 IP 分组封装在以终端 MAC 地址为源 MAC 地址,虚拟 MAC 地址为目的 MAC 地址的 MAC 帧中,只有主路由器对封装在这样 MAC 帧中的 IP 分组进行转发操作,其他 VRRP 路由器即使接收到该 MAC 帧,也将丢弃该 MAC 帧。

7. 交换机转发表更新过程

如果将图 5.27 中以太网扩展为图 5.31 所示以太网结构,在路由器 R2 成为主路由器后,以太网中各个交换机的转发表需要生成表 5.5 所示的转发项,否则,可能导致终端发送给默认网关的 MAC 帧在以太网中广播的情况。为了在各个交换机中生成表 5.5 所示的转发项,当路由器 R2 成为主路由器时,立即发送一个 VRRP 报文,该 VRRP 报文最终被封装成以虚拟 MAC 地址 00-00-5E-00-01-02 为源 MAC 地址,以组地址 01-00-5E-00-00-12 为目的 MAC 地址的 MAC 帧,该 MAC 帧在以太网中广播,如图 5.31 所示,以太网中所有交换机都接收到该 MAC 帧,通过地址学习,在转发表中建立表 5.5 所示的转发项。路由器 R2 通过定期发送 VRRP 报文定期刷新各个交换机中虚拟 MAC 地址对应的转发项,使得各个交换机一直在转发表中维持该转发项。

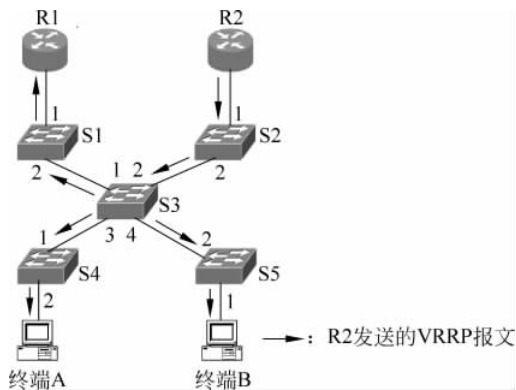


图 5.31 以太网结构

表 5.5 交换机转发表

MAC 地址	转发端口
00-00-5E-00-01-02	交换机 S1
	端口 2
00-00-5E-00-01-02	交换机 S2
	端口 1
00-00-5E-00-01-02	交换机 S3
	端口 2
00-00-5E-00-01-02	交换机 S4
	端口 1
00-00-5E-00-01-02	交换机 S5
	端口 2

8. 负载均衡

图 5.27 所示的 VRRP 工作环境能够解决容错问题,但无法实现负载均衡,为了实现负载均衡,采用图 5.32 所示的 VRRP 工作环境。创建两个 VRID 分别为 2 和 3 的虚拟路由器,同时将路由器 R1 和路由器 R2 连接以太网的接口分配给两个虚拟路由器,为 VRID 为 2 的虚拟路由器分配虚拟 IP 地址 192.1.1.1,使得路由器 R1 因为是 IP 地址拥有者而自然成为 VRID 为 2 的虚拟路由器中的主路由器。为 VRID 为 3 的虚拟路由器分配虚拟 IP 地址 192.1.1.2,使得路由器 R2 因为是 IP 地址拥有者而自然成为 VRID 为 3 的虚拟路由器中的主路由器。将一半连接在网络 192.1.1.0/24 上的终端(图 5.32 中的终端 A)的默认网关地址配置成 VRID 为 2 的虚拟路由器对应的虚拟 IP 地址 192.1.1.1,将另一半连接在网络 192.1.1.0/24 上的终端(图 5.32 中的终端 B)的默认网关地址配置成 VRID 为 3 的虚拟路由器对应的虚拟 IP 地址 192.1.1.2,这样,连接在网络 192.1.1.0/24 上的终端,一半将路由器 R1 作为默认网关,另一半将路由器 R2 作为默认网关,一旦某个路由器发生故障,另一个路由器将自动作为所有终端的默认网关,既实现了容错,又实现了负载均衡。

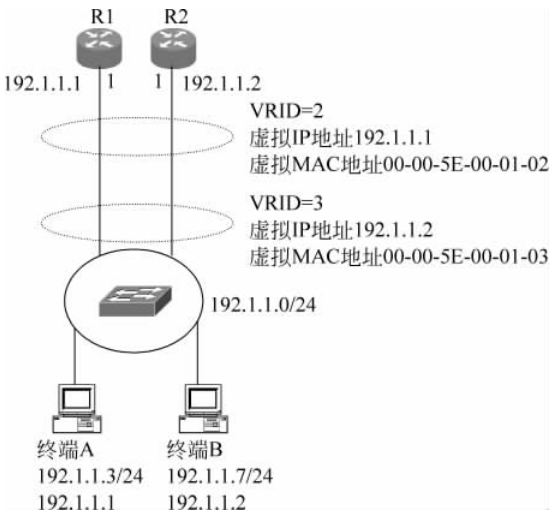


图 5.32 均衡负载的 VRRP 工作环境

5.4.3 VRRP 应用实例

1. 网络结构与基本配置

图 5.33 是图 5.26 的简化版,为了实现容错和负载均衡,对网络进行如下配置。

- 根据图 5.33 所示配置信息分别为路由器 R1 和路由器 R2 的两个接口配置 IP 地址和子网掩码,完成路由器接口 IP 地址和子网掩码配置后,路由器 R1 和路由器 R2 自动生成图 5.33 所示的路由表,路由表中给出用于指明通往路由器直接连接的网络的传输路径的路由项;
- 创建 VRID 分别为 2 和 3 的两个虚拟路由器,并将路由器 R1 接口 1 和路由器 R2 接口 1 分配给 VRID 为 2 的虚拟路由器,并将路由器 R1 接口 2 和路由器 R2 接口 2 分配给 VRID 为 3 的虚拟路由器,VRID 为 2 的虚拟路由器对应的虚拟 MAC 地址为 00-00-5E-00-01-02,VRID 为 3 的虚拟路由器对应的虚拟 MAC 地址为 00-00-5E-00-01-03;
- 为 VRID 为 2 的虚拟路由器分配虚拟 IP 地址 192.1.1.254,这使得路由器 R1 成为 VRID 为 2 的虚拟路由器的主路由器,为 VRID 为 3 的虚拟路由器分配虚拟 IP 地址 192.1.2.253,这使得路由器 R2 成为 VRID 为 3 的虚拟路由器的主路由器;
- 连接在网络 192.1.1.0/24 上的终端配置默认网关地址 192.1.1.254,连接在网络 192.1.2.0/24 上的终端配置默认网关地址 192.1.2.253。

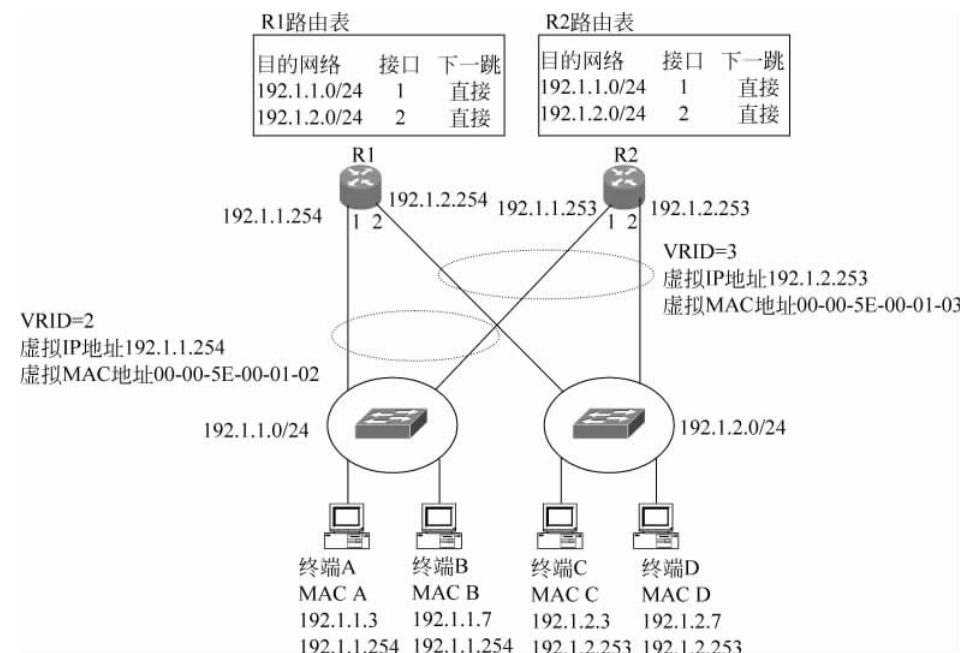


图 5.33 网络结构与基本配置

2. 生成主路由器和转发项

路由器 R1 因为是虚拟 IP 地址 192.1.1.254 的 IP 地址拥有者,自然成为 VRID 为 2 的

虚拟路由器的主路由器,在成为主路由器时,一是通过发送 VRRP 报文,在网络 192.1.1.0/24 各个交换机中建立将目的 MAC 地址为 00-00-5E-00-01-02 的 MAC 帧转发给路由器 R1 接口 1 的转发项。同时,通过在网络 192.1.1.0/24 中广播将虚拟 IP 地址 192.1.1.254 与虚拟 MAC 地址 00-00-5E-00-01-02 绑定的 ARP 报文,在连接在网络 192.1.1.0/24 上的所有终端的 ARP 缓存中建立 IP 地址 192.1.1.254 与 MAC 地址 00-00-5E-00-01-02 的绑定。同样的,在路由器 R2 成为 VRID 为 3 的虚拟路由器的主路由器后,在网络 192.1.2.0/24 各个交换机中建立将目的 MAC 地址为 00-00-5E-00-01-03 的 MAC 帧转发给路由器 R2 接口 2 的转发项。在连接在网络 192.1.2.0/24 上的所有终端的 ARP 缓存中建立 IP 地址 192.1.2.253 与 MAC 地址 00-00-5E-00-01-03 的绑定。

3. IP 分组传输过程

如果终端 A 需要向终端 D 发送 IP 分组,首先获取终端 D 的 IP 地址 192.1.2.7,构建源 IP 地址为 192.1.1.3、目的 IP 地址为 192.1.2.7 的 IP 分组。通过判别终端 A 和终端 D 所在网络的网络地址(192.1.1.0/24 和 192.1.2.0/24)发现终端 A 和终端 D 不在同一个网络,终端 A 需要将 IP 分组发送给默认网关。终端 A 从 ARP 缓存中获取默认网关地址 192.1.1.254 绑定的 MAC 地址 00-00-5E-00-01-02,构建以终端 A 的 MAC 地址 MAC A 为源 MAC 地址,以 MAC 地址 00-00-5E-00-01-02 为目的 MAC 地址的 MAC 帧,网络地址 192.1.1.0/24 保证将该 MAC 帧转发给路由器 R1,路由器 R1 由于是 VRID 为 2 的虚拟路由器的主路由器,必须对封装在以虚拟 MAC 地址 00-00-5E-00-01-02 为目的 MAC 地址的 MAC 帧中的 IP 分组进行转发操作。路由器 R1 从 MAC 帧中分离出 IP 分组,用 IP 分组的目的 IP 地址 192.1.2.7 匹配路由器 R1 路由表中的路由项,发现和路由项 <192.1.2.0/24,2,直接>匹配,下一跳为直接表明目的终端连接在接口 2 连接的网络上,通过 ARP 地址解析过程获取与目的 IP 地址 192.1.2.7 绑定的 MAC 地址 MAC D,构建以接口 2 所属的虚拟路由器对应的虚拟 MAC 地址 00-00-5E-00-01-03 为源 MAC 地址,以终端 D 的 MAC 地址 MAC D 为目的 MAC 地址的 MAC 帧,通过网络地址 192.1.2.0/24 将该 MAC 帧转发给终端 D,终端 D 从 MAC 帧中分离出 IP 分组,完成终端 A 至终端 D IP 分组的传输过程。

当终端 D 向终端 A 发送 IP 分组时,终端 D 先将 IP 分组转发给默认网关——路由器 R2,实现了路由器 R1 和路由器 R2 的负载均衡。当其中一个路由器发生故障,另一个路由器将作为连接在两个网络上的终端的默认网关。

习题

- 5.1 为什么说 IP 是一种网际协议? IP 实现连接在不同传输网络上的终端之间通信的技术基础是什么?
- 5.2 为什么为每一个路由器接口分配 IP 地址?
- 5.3 作为中继系统,转发器、网桥和路由器有何区别?
- 5.4 解释不能用网桥实现两个分别连接在以太网和 ATM 网络的终端之间通信的原因。

网络地址配置,给出路由器 R1、路由器 R2 和路由器 R3 的路由表。

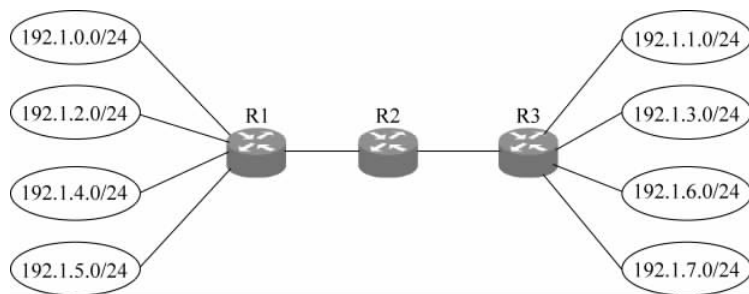


图 5.35 题 5.29 图

5.30 根据图 5.36 所示的互连网络结构,为每一个局域网分配合适的网络前缀地址(假定 CIDR 地址块为 192.77.33.0/24,图中每一个局域网旁边标明的数字是该局域网的主机数)。

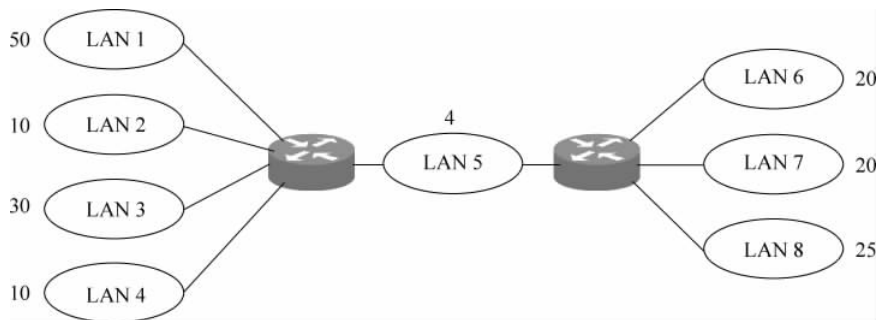


图 5.36 题 5.30 图

5.31 某单位分配到一个地址块 136.23.12.64/26,现在需要进一步划分为 4 个一样大的子网,试问:

- (1) 每个子网的网络前缀有多长?
- (2) 每个子网有多少地址?
- (3) 每个子网的地址块是什么?
- (4) 每个子网可分配给主机的最小和最大地址是什么?

5.32 网络结构如图 5.37 所示,给出的 CIDR 地址块是 192.1.1.64/26,确定每一个子网的网络地址,将最大可用地址分配给路由器连接对应子网的接口,给出路由器 R1、R2 的路由表。

5.33 互连网络结构如图 5.38 所示。

① 补齐图中终端和路由器的配置信息,包括路由表。使其能够实现终端 A 和终端 B 之间的 IP 分组传输。

② 以①补齐的配置信息为基础,给出终端 A 至终端 B IP 分组传输过程中涉及的所有 MAC 帧,并给出这些 MAC 帧的源和目的 MAC 地址(假定终端和路由器的 ARP 缓冲器为空)。

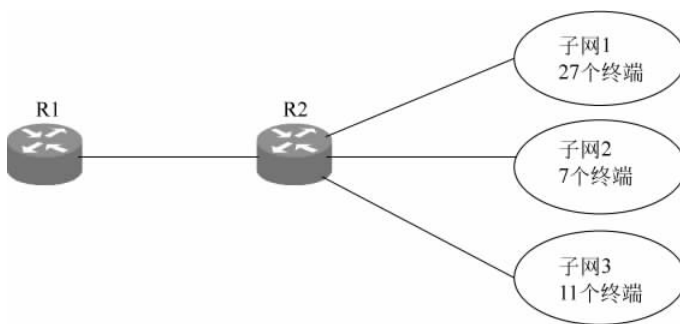


图 5.37 题 5.32 图

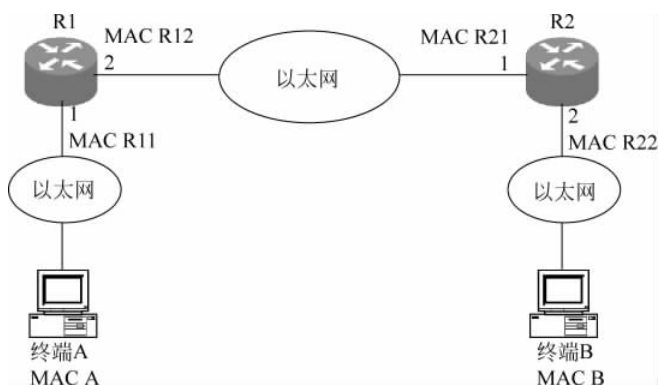


图 5.38 题 5.33 图

5.34 VRRP 的作用是什么？

5.35 简述主路由器转换为备份路由器的条件。

5.36 简述备份路由器转换为主路由器的条件。

5.37 对于图 5.33 所示网络结构,如果要求连接在网络 192.1.1.0/24 和 192.1.2.0/24 上终端,各有一半以路由器 R1、路由器 R2 为默认网关,给出实现这一功能所需的 VRRP 配置。